AFRL-RY-WP-TR-2015-0041

UNCERTAINTY PROPAGATION AND THE FANO BASED INFORMATION THEORETIC METHOD
A Radar Example

John A. Malas and Patricia A. Ryan

Radio Frequency Exploitation Branch
Layered Sensing Exploitation Division

John A. Cortese

Massachusetts Institute of Technology, Lincoln Laboratory

FEBRUARY 2015
Final Report

Approved for public release; distribution unlimited.

See additional restrictions described on inside pages

STINFO COPY

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the USAF 88th Air Base Wing (88 ABW) Public Affairs Office (PAO) and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2015-0041 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

JOHN MALAS
Program Manager, Lead Researcher
RF Exploitation Branch
Layered Sensing Exploitation Division

//Signature//

JEFFREY SANDERS, Chief
RF Exploitation Branch
Layered Sensing Exploitation Division

//Signature//

JACQUILINE BARKER
Division Chief
Layered Sensing Exploitation Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YY)***
February 2015

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
11 January 2010 – 26 October 2012

4. TITLE AND SUBTITLE
UNCERTAINTY PROPAGATION AND THE FANO BASED INFORMATION THEORETIC METHOD

5a. CONTRACT NUMBER
In-house

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)
John A. Malas and Patricia A. Ryan, (AFRL/RYAP)
John A. Cortese, MIT Lincoln Laboratory

5d. PROJECT NUMBER
10RY08C0R

5e. TASK NUMBER
N/A

5f. WORK UNIT NUMBER
N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Radio Frequency Exploitation Branch
Layered Sensing Exploitation Division
Air Force Research Laboratory, Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command, United States Air Force

Massachusetts Institute of Technology
Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-RY-WP-TR-2015-0041

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

Air Force Office of Scientific Research
AFOSR/RSE
875 North Randolph Street, Suite 325, Room 3112
Arlington, VA 22203-1768

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RYAP

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RY-WP-TR-2015-0041

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
The U.S. Government is joint author of the work and has the right to use, modify, reproduce, release, perform, display or disclose the work. PAO case number 88ABW-2013-0274, Clearance Date 23 Jan 2013. Report contains color.

14. ABSTRACT
This report contains work performed under AFOSR Lab Task 10RY08C0R. The Fano equality is joined with the data-processing inequality to develop a theory model for component level trade studies within radar signature exploitation systems. Entropy is used to represent propagating uncertainty within an information channel. Measures are developed to identify information flow bottlenecks within an information loss budget. The propagating effects of various sources of uncertainty on system performance are characterized. Numerical computation of entropic estimates on high dimensional signature processes are explored for risk-based design methods. Methods are demonstrated on a radar example.

15. SUBJECT TERMS
information theory, radar, uncertainty, systems modeling, information sensing, target recognition

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
Unclassified	Unclassified	Unclassified	SAR	76	John Malas
					19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Table of Contents

1. INTRODUCTION

Innovations in sensing component technology are spurring new research in the area of target signature measurement and exploitation. Innovations include multi-channel spatially diverse antennas, sensitive receivers, fast analog-to-digital converters, adaptive transmit waveforms, and sparse sampling approaches. These innovations support new signature information sensing functions including calibrated target measurements, feature processing, and inference based decision algorithms. The ability to characterize *information* extraction while under the effects of system uncertainties is critical to risk based design methods. The use of existing systems theory prototypes such as the radar range equation is inadequate to fully characterize the flow of information through the sensing system.

The success of any theory model in the above context will largely depend on its ability to address several challenges; (1) ability to characterize information gain and performance within various stages of the system, (2) propagate the effects of these uncertainty sources acting on individual components within the system to the predicted *system* performance measures, (3) effectively minimize the overall loss in the information flow while trading costs associated with component design, and (4) operate effectively within the nonlinear high dimensional spaces inherent in signature sensor systems.

1.1 Case for Information Theoretic

The use of information theoretic principles affords several advantages in dealing with the challenges in the information sensing and exploitation areas. First, information theory prototypes enable the study of the propagating effects of uncertainty on system performance at the various points of noise infiltration. Using Fano's inequality, the max flow [1] criteria bounds the optimal Bayes error. Entropy and mutual information (MI) are analytically connected to the probability of error (P_e) and more generally the Neyman Pearson criteria allowing for the rate of noise infiltration to be related to the rate of entropy growth and ultimately to the rate of degradation system performance. The information loss associated with uncertainty sources can then be characterized in terms of a confidence interval about the predicted system performance at each component of the system. The data processing inequality affords a method to determine information loss points and maximize information flow via *component* trades within a system information loss budget.

Second, the convexity of mutual information yields a unique solution and enables rapid numerical convergence (low computational complexity) to maximum MI configurations [2], [3]. MI affords the optimization of a scalar quantity while classical Bayes likelihood ratio techniques can involve optimizing on non-convex surfaces over high dimensional signature spaces. On a convex surface, the use of highly efficient search algorithms such as the Conjugate Gradient method will converge on the order of N operations (N dimensional problem). While entropy based methods operate non-parametrically such that the probability does not have to be estimated, complicating factors can include numerical computation issues that occur within high dimensional processes (Bellman's Curse of Dimensionality). It can be shown however [4] that computing the entropy of the multivariate sensor signature processes is also $O(N)$. As a consequence of the law of large numbers, the asymptotic equipartition property asserts that there are large regions within the entropic signature subspace which will never occur under the decision hypotheses [2]. Thus, the information theoretic approach holds the potential to exploit entropy based methods operating within this "typical" signature subspace.

Third, classical statistical pattern recognition approaches use the maximum likelihood (ML) decision criteria which include only the 2nd order statistics present in the training process. The use of MI in nonlinear processing affords advantages over linear processing in that it accounts for higher-order statistics within the design of nonlinear optimal decision rules and in the optimization of features. Nonlinear scattering phenomenon resulting from the interaction of individual target mechanisms can also reduce the effectiveness of second order techniques [5] in the optimization of diverse transmit waveforms. The use of MI as a nonlinear signal processing method for optimizing waveform design will address this phenomenon. It is these inherent benefits that distinguish the information theoretic approach over traditional statistical pattern recognition methods.

1.2 Historical Contributions

The use of information theory in the area of radar dates back to the early 1950s. Woodward and Davies [6] and Woodward [7] were the first to apply the information theoretic approach to the analysis of radar soon after the appearance of Shannon's original work [8] on information theory. More recently Bell [9] has suggested the use of an information theoretic approach to the design of radar waveforms. Dr. Bell formulated and obtained a solution to the problem of designing a waveform that maximized the MI between the target impulse response (viewed as a random process) and the received signal. Recently, Leshem et al. [10] extended Bell's work to the case of multiple extended targets. Sowelam and Tewfik [11] also used waveform design in conjunction with the Kullback-Liebler [12] criterion to distinguish between different target classes. Briles [13] applied rate distortion theory to analyze impulse radar for use in target identification design and performance prediction. Horne and Malvern [14] introduce a high level theoretical framework to calculate the information conveyed by the image of a target based on pixel values relative to the modeled fluctuations of these values. Principe, Xu, Zhao, and Fisher [15] present a framework for learning based on information theoretic criteria and have studied the application of the Fano bound to the ATR problem set. Methods such as the maximum likelihood test have been used to evaluate radar signature processes for target classification performance as in the work by O'Sullivan et al [16]. This framework proposes several approximations to the Kulback-Leibler divergence that can be used to estimate statistical distances compatible with pattern matching algorithms. Pasala and Malas [17] introduce the use of MI as a similarity measure for use in the evaluation of suitability of radar signature training surrogates. Recently, there has been much interest in radars with a new architecture referred to as the MIMO (Multiple Input Multiple Output) radar [18] – [21]. It is the information theoretic approach that unifies the analysis of these radar systems.

Several contributions within the body of existing referenced work including [9] and [13] have (in one form or another) presented the radar system in terms of a Markov Chain within a channel configuration and characterized the information flow from source to sink. Tishby [22] has developed the information bottleneck approach, wherein rate-distortion theory, the Data Processing Theorem, and compression play major roles. The max-flow min-cut application to the channel problem has been studied to understand the relationship of capacity to information flow [23]. Ahlswede and Yeung [1] have extended this analysis to network information flow where a single information source is multicast to multiple destinations. B. C. Geiger et-al [24] have studied the information loss induced by static nonlinearities within a memoryless nonlinear input-output system and conclude that a particular output can result from multiple inputs. Merhav [25] have published a series of works on information measures with application to estimation theory. While this work is thematically similar to the work herein, the author has used a different theoretical approach to the study of information theory and decision theory.

The area of uncertainty modeling and sensitivity analysis is wide ranging drawing upon the established fields such as the design of experiments [26] and classical engineering methods of statistics that lead to uncertainty measures [27]. The subject of propagation of uncertainty has been firmly established within traditional methods of Taylor Series expansion and differential calculus [28]. Recently, advances in large scale computer simulation have opened the door to modeling complex physical processes in lieu of expensive physical experiments [29]. The methods by Saltelli et al. [30] present new Monte-Carlo Methods for the study of sensitivity.

A significant contribution of the research reported to the existing body of work is the development and demonstration of a systems theory model for the study of the effects of uncertainty on the information flow within the various *components* of the sensor system. Fano's equality is developed in a mathematically concise form. The Fano's equality is joined with the data-processing inequality to address nonlinear paradigms in sensing. While not new n the literature, the unified use of the Fano equality with the Data Processing inequality in a Markovian channel construct is a center piece of this work. An abbreviated and preliminary treatment of this concept is presented in [31]. Entropy is used to model propagating uncertainty within an information channel. Measures are developed to identify information flow bottlenecks. The mathematical framework for an information theoretic approach to estimation and hypothesis testing is applied to a multidisciplinary problem set. Techniques for bounding asymptotic performance under sufficient statistics are characterized and related to phase transitions within the typical set trajectory associated with sampling uncertainty. Nonparametric performance estimations are developed at various points in the information sensing pipeline. Minimum sampling requirements for performance prediction are developed on high dimensional signature processes based on low sample entropic estimates.

2. Approach

2.1 Systems Theory Model

Taking an information theoretic view, degrading effects are considered as sources of entropy. Treating the system as an information flow pipeline from input to output, the injective entropy acts to degrade the Shannon MI between the input and output. The systems model is demonstrated as a suitable vehicle for performing component level design trades within the information sensing application based on a component level information loss budget (Bits). Demonstration of the max flow in conjunction with the Data Processing Inequality provides analysis of "bottlenecks" in the information flow pipeline. Key attributes of the theoretical model have been demonstrated under the constraints of a radar high range resolution sensor system example. Modeling and simulation for simplified target scattering models are used to illustrate the value of component level analysis under the propagating effects of various sources of uncertainty.

2.2 Addressing Dimensionality

Interdependencies among multivariate target signatures can significantly impede information extraction. The number of samples required to estimate the underlying signature statistics is related to the incremental increases in uncertainty. Baseline statistical sample requirements (in the native coordinate system) associated with the resolved radio frequency target scattering are characterized for specified states of certainty. Methods are developed that estimate the sampling requirements for entropic quantities based on a characterization of the typical set underlying the sufficient statistics of the random signature process. The variance of the performance estimate associated with low sample count Monte-Carlo experiments can be scaled (via central limit theorem) to the estimate the performance variance associated with higher sample counts.

2.3 Uncertainty Analysis

Both sensor uncertainty and model training uncertainty are propagated into the classifier where uncertain decisions are inferred from uncertain observations. The uncertainty (increase in entropy) is ultimately realized in the form of confidence or *reliability* intervals about the estimated system performance. Mathematically defined categories of uncertainty are developed to better understand the entropic effects within the information sensing system. A sensitivity analysis is performed to study the relative significance of various “unknown” operating conditions to the reliability of the performance estimate at each component of the system. The effects of sampling uncertainty are contrasted to reliability of performance estimates. This comparison forms the basis for the study of the variance effects in performance estimation within high dimensional signature processes subject to unknown operating conditions.

3. THEORY

3.1 Uncertainty Sources within an Information Sensing System and the Decision Rule Subspace

It is important to contrast the proposed concept of uncertainty with several terms generally used by the measurement community [32]. Accuracy refers to the agreement between a measurement and the true or correct value. Precision refers to the repeatability of a measurement. Error refers to the disagreement between a measurement and the true or accepted value. The uncertainty in a stated measurement is the interval of confidence around the measured value such that the measured value is expected not to lie outside this stated interval. The use of the term "uncertainty" implies that the true value may not be known and can be stated along with a probability. These definitions recognize the deterministic nature of error and the stochastic nature of uncertainty. However, uncertainty as currently defined by the sensor measurement community may not be sufficient to address the full range of issues under study within an information sensing system.

Modern sensing systems produce signature measurements that when combined with the effects of various system uncertainties are realized as a random signature process. Conclusions are inferred by applying instances taken from this random measured signature process to a decision rule. The "unknowable" nature of parameters affecting the measured signature process leads to challenges in developing a signature process model that will generate an optimal decision rule for inferring information. The combined effects of these sensing and training uncertainties limit the exploitation of physics-based features and result in a loss in information that can be extracted from target signature measurements. The resulting decision uncertainty is then driven by both the distorted measurements and the degree of agreement between the signature process under measurement and the process used to train the optimal decision rule.

3.2 Sensing Uncertainty Example

It is helpful to illustrate the concepts surrounding uncertainty through the use of a real world example. A related problem of interest is the measurement of airborne moving objects using high range resolution (HRR) waveforms. The successful extraction of information from these measurements via a specific system design is complicated by several sources of uncertainty. The two general classes of system uncertainty introduced above are given in Table I as "Sensing" uncertainty and uncertainty resulting from "Decision Rule Training Limitations". Sensing uncertainty is divided into three subcategories (a) signature measurement uncertainty due to sensor design/limitations, (b) the uncertainty due to interference, and (c) object tracking position and motion uncertainty.

The object under measurement by the sensing radar system can be viewed as a collection of scattered field sources filling an electrically large volume in space. The system measurement of this object is subject to uncertainty identified in source 1.a generating the statistical support underlying a random signature process at a fixed position in time. Target fixed body motion within the measurement interval induces scintillation within the scattering sources resulting in an additional increase in entropy. Imperfect knowledge of target position, velocity, and aspect also alters the statistical characterization of the random signature process (source 1.c). The random signature process also interacts with an external environment (source 1.b) to further impact the statistical nature of the measured signature process.

TABLE 1
RADAR SYSTEM UNCERTAINTY SOURCES

Uncertainty Core Area	*Parameter Uncertainty Subcategory*		
1. Sensing			
a. Signature Measurement*	Non-linear Effects	I&Q Quantization/ Clipping	Amplitude & Phase Calibration
b.Environmental	Clutter/Thermal Noise	RF Interference	Jamming
c. Object Tracking & Motion	Object Range, Velocity, & Aspect Estimates	Object Articulation	Intra-measurement Motion
2. Decision Rule Training Limitations	Process Under Sampling*	Target Configuration Variation	Target Modeling Parameters

* Epistemic Uncertainty

The exploitation of this signature process using a decision algorithm requires the training[1] of an optimal decision rule that operates within the entropy produced by sources 1.a, 1.b, and 1.c. Only a subset of the phenomenon (parameters) underlying source 1 can be modeled and/or characterized within the statistical decision rule training process. Limitations within the training process result in a decision rule design that is less than optimal with respect to system performance.

Sources of uncertainty that arise because of natural, unpredictable variation within the system under are aleatoric and are considered "unknowable". Source 1 uncertainties are aleatoric in nature and as such can only be characterized statistically. Uncertainties that are conceptually resolvable yet subject to systematic limitations are epistemic and are to be reduced as much as possible within the analysis. The performance resulting solely from signature measurement uncertainty source 1.a (highest system certainty state) is defined here as epistemic in nature and in general can be modeled and characterized. Sources 1.b, 1.c, and 2 are in general aleatoric and can result in a reduction in certainty from the highest certainty state.

[1] Supervised learning assumed as the classification training approach.

3.3 Uncertainty and the Decision Rule Subspace

The sources of uncertainty associated with source 2 in Table I are traceable to the corresponding effects within the decision rule subspace in the classical statistical pattern recognition approach to the binary hypothesis testing. The decision rule design (decision threshold d) is based on the statistical training support resulting from the uncertainties in Table I. If the sensing uncertainties within source 1 are adequately represented in the statistics of the training process, the decision rule design should provide optimal performance. The effects due to many of the uncertainties in Table I are unavoidable. For example, target signature realizations are often formed through the integration of many sequential measurements. Intra-measurement object motion can cause distortion and induce uncertainty in the decision rule subspace that is not accounted for in the decision rule training process. In another example, the object under measurement may be configured differently than that represented in the training data (extra fuel tanks, wing flaps up, or damaged surface for example).

3.4 Information Theoretic Decision Rule Subspace

An alternative to the classic statistical pattern recognition approach to viewing the decision rule subspace is shown in Fig. 1. In Fig. 1, the decision rule subspace is cast in terms of information theoretic quantities based on entropy; a measure of the size of a typical set [2]. In Fig. 1, information is defined in terms of the MI between the "typical subspaces" [2] associated with the true object state **H** and the decision state **Q** where **H** and **Q** are discrete random variables. Systems (and associated sub-component) designs that increase the MI between these "typical signature subspaces" increase the flow of information. "Uncertainty" acts to alter to the typical signature subspaces (growth or movement) associated with the highest certainty state. A change to the typical subspaces can result in a loss in the flow of information and a decrease in decision performance.

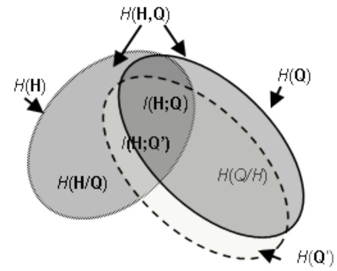


Fig. 1. Decision Rule Subspace and the Overlapping Typical Sets

*Modified version of Figure 2.2, [2]

3.5 Information Theoretic Radar Channel Model

The concept of uncertainty introduced in Fig. 1 can be realized in terms of an increase in entropy within a discreet memoryless information channel. The radar information sensing system can be viewed within this channel model depicting the information flow through the signature sensing and processing components of a radar system as shown in Fig. 2 [2], [33].

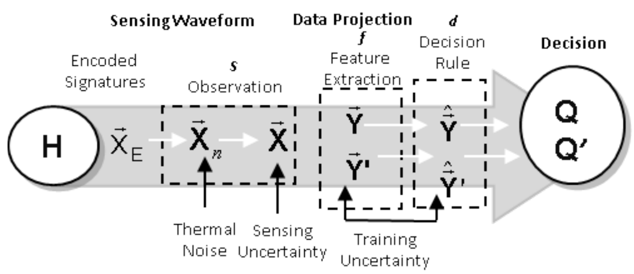


Fig. 2. Information Theoretic Sensor Channel Model

The relationship between **H** and **Q** is the basis chosen for performance characterization. The discrete random variable **H** represents which of the N_c possible hypotheses has occurred. **Q** is of the same alphabet as **H.** Successful flow of information results in agreement between **H** and **Q**.

Conditioned on the generating hypothesis H (instance of **H**), there is typically a multidimensional encoded source[2] $\vec{X}_E$ which is realized as the image projection of the scattered field of the object under measurement. In the case of an HRR radar measurement, this is the band-limited frequency response associated with the scattered field of the observed object in thermal noise. After mixing, filtering, and signal processing, these returns become the measured random signature vector $\vec{\mathbf{X}}_n$. The sensing of $\vec{\mathbf{X}}_n$ is subjected to the uncertainties listed in source 1 of Table I leading to the random signature process $\vec{\mathbf{X}}$. The various cases of the sensed signature are summarized in Table II.

TABLE II

SENSOR MEASUREMENT SIGNAL CASES

Encoded Deterministic Multivariate Signal	$\vec{X}_E$
Deterministic Signal in Additive Noise	$\vec{\mathbf{X}}_n = \vec{X}_E + \vec{\mathbf{n}}$
Random Multivariate Signal in Additive Noise	$\vec{\mathbf{X}} = \vec{\mathbf{X}}_E + \vec{\mathbf{n}}$

The multivariate sample feature $\vec{Y}^i$ is extracted from the i^{th} instance test sample $\vec{X}^i$ to support the desired function of the exploitation system. Given the random nature of $\vec{\mathbf{X}}$, the extracted signature feature $\vec{\mathbf{Y}}$ is also random. The training feature process $\vec{\mathbf{Y}}'$ is developed from the set of typical signatures within the decision rule training process $\vec{\mathbf{X}}'$. $\vec{\mathbf{X}}'$ (and thus $\vec{\mathbf{Y}}'$) is developed offline using a surrogate process and is used to determine the 'optimal' decision rule *d*. The decision algorithm applies $\vec{Y}^i$ to the decision rule *d* yielding the decision Q (instance of **Q**) declaring which of the hypothesis has occurred. The evaluation of an ensemble N_M of test samples $\vec{X}^i$ $\{i=1 \rightarrow N_M\}$ produces the sample ensemble of the N_M matching tests (H,Q) to statistically characterize the decision performance.

[2] $\vec{X}_E$ in this context is deterministic resulting from the convolution of the target's physical scattering mechanisms with the transmitted waveform *s*. Given the "unknowable" nature of this code through measurement or modeling, the code itself is only observable in the random form of $\vec{\mathbf{X}}$.

3.6 Fano Based Information Theoretic Method (FBIT)

It is desired to quantify the effects of "uncertainty" and the associated alteration to the typical signature subspaces in terms of the flow of information and the impact to system performance. Two theorems from information theory play key roles in the development of these relationships. Fano's Inequality relates information theoretic quantities to the P_e criterion for an object classification system [2]. The Data Processing Inequality [2] affords the analysis of the flow of information from measured object returns through the signature sensing, signal processing architecture, and into the decision stage; detailing where information is lost, and quantifying the impact on system performance. In this manner, stages in the information processing pipeline where information is lost can be identified, analyzed and optimized, leading to improvement in overall system performance.

Fano's inequality provides a mathematical means to relate the MI between **H** and **Q**, $I(\mathbf{H};\mathbf{Q})$, to a lower bound on P_e. Fano's inequality [2] can be written as an equality as in (1).

$$H(P_e) = \delta - P_e \cdot \log(N_c-1) + H(\mathbf{H}/\mathbf{Q}) \tag{1}$$

In (1), P_e is a real random variable between 0 and .5 representing the probability of error of the decision algorithm. N_c is the discrete size of the alphabet of **H** and **Q.** $H(\mathbf{H})$ is the Shannon entropy of the discrete random variable **H**. δ is a bias offset derived from asymmetries in the data and decision algorithm [32]. Typically δ is small and to a first approximation may be neglected.

Theorem I: For $N_c = 2$, Fano's equality can be written as $H(P_e) = 1 - I(\mathbf{H};\mathbf{Q}) + I(\mathbf{Q};\mathbf{V})$ where **V** is the binary discrete random variable representing the probability that the decision rule makes a correct decision.

Proof Sketch: Using $I(\mathbf{H};\mathbf{Q}) = H(\mathbf{H}) - H(\mathbf{H}/\mathbf{Q})$ and (1) we get (2) below.

$$H(P_e) = \delta - \mathbf{P_e} \cdot \log(N_c - 1) + H(\mathbf{H}) - I(\mathbf{H;Q}) \qquad (2)$$

The asymmetry factor in (2) can be computed directly from the output of the decision algorithm. Let $\delta = I(\boldsymbol{Q;V})$ for $N_c = 2$; where $\mathbf{V}$ is the binary discrete random variable representing the probability that the decision rule makes a correct decision. $\mathbf{V} = 1$ when $\mathbf{H} = \mathbf{Q}$; otherwise $\mathbf{V} = 0$. Equation (2) can then be written more completely for $N_c = 2$ as in (3) below [32].

$$H(P_e) = 1 - I(\mathbf{H;Q}) + I(\mathbf{Q;V}) \qquad (3)$$

Equation (3) can be written in terms of the inverse entropy function F as shown in (4).

$$P_e = F(H(\mathbf{H}) - I(\mathbf{H;Q}) + I(\mathbf{Q;V})) \qquad (4)$$

F is a deterministic strictly monotonically increasing function that maps information theoretic quantities into the P_e at the corresponding operating point. The relationship of P_e to $F(x)$ where $x \in [0, ½]$ is given Fig. 3.

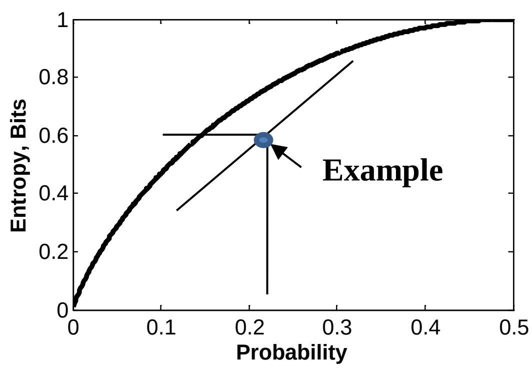


Fig. 3. The Binary Entropy Function, Bits

The quantity IL_Q in (5) is the end to end information loss for the system.

$$\mathsf{IL}_{\mathbf{Q}} = \mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H;Q}) \quad (5)$$

Minimizing the information loss minimizes the system P_e.

The entropic quantity H(**H**) is determined by the a priori probabilities of the outcomes of the random variable **H** corresponding to the different target classes. δ is fixed by architectural considerations. Since F is a known function, the deterministic relation $\mathsf{P}_\mathsf{e} = \mathrm{F}(\mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H;Q}) + \mathrm{I}(\mathbf{Q;V}))$, for fixed H(**H**) and δ, determines the MI (I(**H;Q**)) needed to achieve a specified P_e. For example, for an equiprobable binary hypothesis scenario[3], H(**H**) = 1 Bit and I(**Q;V**) $\approx$ 0, an approximation for P_e can be written as

$$\mathsf{P}_\mathsf{e} \approx \mathrm{F}(1 - \mathrm{I}(\mathbf{H;Q})). \quad (6)$$

Specifying a desired P_e determines the amount of allowed IL_Q. How the IL_Q budget is "spent" as information cascades from the input space at **H** to the classifier output space at **Q** can be traded off via component (link) design. Fig. 4 presents an abstract diagram indicating possible tradeoffs. Information losses within the channel can be studied with respect to various sources of uncertainty in Table I.

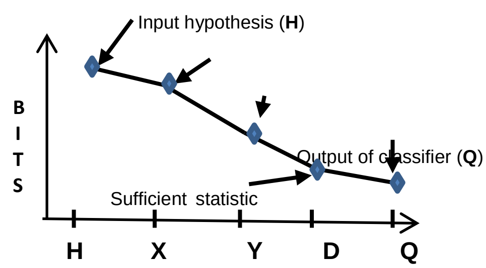


Fig. 4. Example of Trading System Component Level

Design for Information in Bits [33]

[3] The selection of the uniform prior on **H** is for illustration purposes and without loss of generality.

The Data Processing Inequality states that information can only be lost in the channel as shown in (7).

$$I(\mathbf{H};\vec{\mathbf{X}}) \geq I(\mathbf{H};\vec{\mathbf{Y}}) \geq I(\mathbf{H};\mathbf{Q}) \tag{7}$$

Using the relationship in (5) and (4) the loss associated with each link within the channel can be characterized as in (8).

$$H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{X}}) \leq H(\mathbf{H}) - I(\boldsymbol{H};\vec{\mathbf{Y}}) \leq H(\mathbf{H}) - I(\mathbf{H};\boldsymbol{Q}). \tag{8}$$

The *cumulative* information loss at each link in the channel can then be written as below applying (5).

$$\mathrm{IL}_{\vec{X}} = H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{X}}); \quad \vec{X} \in |\chi| \tag{9.a}$$

$$\mathrm{IL}_{\vec{\mathbf{Y}}} = H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{Y}}); \quad \vec{Y} \in |Y| \tag{9.b}$$

$$\mathrm{IL}_{\mathbf{Q}} = H(\mathbf{H}) - I(\mathbf{H};\mathbf{Q}); \tag{9c}$$

The respective information loss due to each link within the Markov chain $\mathbf{H} \rightarrow \mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Q}$ can then be defined as in equations (10a-10c).

$$\text{Loss due to Sensing} \equiv \mathrm{IL}_{S\Delta} = H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{X}}) \tag{10.a}$$

$$\text{Loss due to Feature Extraction} \equiv \mathrm{IL}_{F\Delta} = I(\mathbf{H};\vec{\mathbf{X}}) - I(\mathbf{H};\vec{Y}) \tag{10.b}$$

$$\text{Loss due to Decision Rule} \equiv \mathrm{IL}_{D\Delta} = I(\mathbf{H};\vec{Y}) - I(\mathbf{H};\mathbf{Q}) \tag{10.c}$$

Thus, the probability of error can be estimated at various points in the channel using the approximation in (6);

$$P_e^X \approx F(H(\mathbf{H}) - I(\mathbf{H};\vec{X})) \tag{11.a}$$

$$P_e^Y \approx F(H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{Y}})) \tag{11.b}$$

$$P_e^Q \approx F(H(\mathbf{H}) - I(\mathbf{H;Q})). \tag{11.c}$$

3.7 Uncertainty in the Information Channel

From (2) we see that sources of uncertainty introduced in the channel can result in a reduction in $I(\mathbf{H;Q})$ and subsequently an increase in P_e. A decrease in $I(\mathbf{H;Q})$ is always accompanied by an increase in $H(P_e)$ resulting in a degradation to the realized P_e.

3.7.1 Categories of Uncertainty

It is important to distinguish between the two distinct categories of uncertainty in Table I. The source 1 uncertainty results from sensing while the source 2 uncertainty results from decision rule training limitations

3.7.1.1 Loss Due to Sensor Process

The loss at $\vec{\mathbf{x}}$, $IL_{S\Delta}$, is due solely to the sensing process. The sensing uncertainty inherently alters the statistics associated with $\vec{\mathbf{x}}_n$ generating statistical independence between $\vec{\mathbf{x}}_n$ and $\vec{\mathbf{X}}$ thus degrading the performance of the signature sensing process as quantified by P_e^X in (11a). The loss in information due to the sensor uncertainty is then realized at $\vec{\mathbf{x}}$ as $IL_{S\Delta}$ in (11a) and is quantified by the entropy $H(P_e^X)$.

$$H(P_e^X) \approx H(\mathbf{H}) - I(\mathbf{H};\vec{\mathbf{X}}) \tag{12}$$

3.7.1.2 Decision Uncertainty Loss

The level of statistical agreement between $\vec{\mathbf{x}}$ and $\vec{\mathbf{X}}'$ will directly affect the loss in the channel due solely to the decision process. The sensing uncertainty sources in Table I are to some degree reproducible in the decision rule training process $\vec{\mathbf{X}}'$. However, sources 1 (b) and 1 (c) in Table I are not fully reproducible in $\vec{\mathbf{X}}'$. The dissimilarity between $\vec{\mathbf{X}}$ and $\vec{\mathbf{X}}'$ results in a decision rule *d* that is not optimal. The application of *d* to the feature process $\vec{\mathbf{Y}}$ induces a loss in the channel due to imperfect training.

The effects of decision uncertainty within the decision rule subspace are realized at **Q** as $IL_{D\Delta}$ as illustrated in Fig. 2. The decision uncertainty $IL_{D\Delta}$ can be interpreted in terms of the entropy $H(P_e^Q)$ as in (13) and quantified as defined in (11.c).

$$H(P_e^Q) \approx H(\mathbf{H};\mathbf{Q}) - H(\mathbf{Q}). \tag{13}$$

3.7.1.3 Training Uncertainty Loss

The feature extraction *f* and decision rule *d* in Fig. 2 are designed to maximize $I(\mathbf{H};\mathbf{Q})$. The resulting $H(\mathbf{P_e^Q})$ provides the best possible performance for a given component design (radar sensor design, feature selection, algorithm design, and decision rule design). As stated above, $\vec{\mathbf{X}}$ is often not completely observable, and a training surrogate $\vec{\mathbf{X}}'$ is used to develop *f* and *d*. Under conditions such as those listed in uncertainty source 2 in Table I, the surrogate representation $\vec{\mathbf{X}}'$ used in the training of the decision rule results in a non-optimal *d*. This is represented by the altered entropic quantity $H(\mathbf{Q'})$ and more importantly $I(\mathbf{H};\mathbf{Q'})$. The alternate Markov chain $\mathbf{H} \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}' \rightarrow \mathbf{Q}'$ is shown as the dotted subspace $H(\mathbf{Q'})$ in Fig. 1. The corresponding form of (3) can then be written as

$$H(\mathbf{P_e}') = 1 - I(\mathbf{H};\mathbf{Q'}) + I(\mathbf{Q'};\mathbf{V})$$

Therefore since $H(\mathbf{P}_e') \geq H(\mathbf{P}_e)$,

$$I(\mathbf{H};\mathbf{Q'}) - I(\mathbf{Q'};\mathbf{V}) \leq I(\mathbf{H};\mathbf{Q}) - I(\mathbf{Q};\mathbf{V}).$$

Corollary I: Information loss due to imperfect training, $IL_{T\Delta}$, is then mathematically quantified in terms of the increase in entropy $\Delta H(\mathbf{P_e})$ resulting from a non-optimal design of *f* and *d*.

$$IL_{T\Delta} = \Delta H(\mathbf{P_e}) = H(\mathbf{P_e'}) - H(\mathbf{P_e}) \quad (14)$$

$$= -I(\mathbf{H};\mathbf{Q'}) + I(\mathbf{Q'};\mathbf{V}) + I(\mathbf{H};\mathbf{Q}) - I(\mathbf{Q};\mathbf{V})$$

If it can be shown that $I(\mathbf{Q};\mathbf{V}) \cong I(\mathbf{Q'};\mathbf{V})$ and that

$$I(\mathbf{Q};\mathbf{V}) << H(\mathbf{H}) - I(\mathbf{H};\mathbf{Q}) \text{ and } I(\mathbf{Q'};\mathbf{V}) << H(\mathbf{H}) - I(\mathbf{H};\mathbf{Q'})$$

then;

$$\text{Imperfect Training Loss} \equiv IL_{T\Delta} \cong I(\mathbf{H};\mathbf{Q}) - I(\mathbf{H};\mathbf{Q'}). \quad (15)$$

The decrease in information flow due to imperfect training is illustrated in Fig. 1 as the reduction in overlap between the subspaces of **H** and **Q**.

Definition I:

The total loss in the channel is equal to the sum of all link loss components.

$$IL_{Total} = IL_{S\Delta} + IL_{F\Delta} + IL_{D\Delta} + IL_{T\Delta} \tag{16}$$

Definition II: Any phenomenon producing a increase in I(**H;Q**) and a subsequent reduction in H($\mathbf{P_e}$) can be defined as a "*system information gain*" within the information channel. Any phenomenon producing a decrease in I(**H;Q**) resulting in an increase in H($\mathbf{P_e}$) is defined as a "*system information loss*".

3.7.2 Propagating Effects of Uncertainty

Uncertainty propagation is the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model inputs [30]. Fig. 5 provides an illustration of a modeling and analysis approach to uncertainty propagation within the sensitivity analysis and modeling of an information sensing system [30].

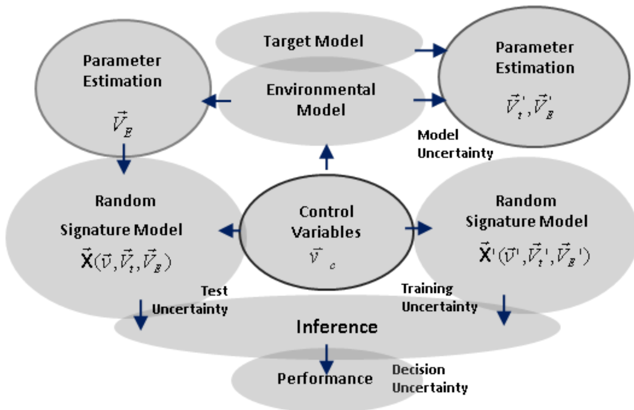


Fig. 5. Parametric Bootstrap Uncertainty and Sensitivity Analysis

The careful definition of variables plays a central role in case controlled studies of the effects of uncertainty on system performance. The vector $\vec{v}_c$ represents the control parameters of interest within computer generated experiments. Absent the uncertainties identified in Table I, the effects of selected values for $\vec{v}_c$ on the deterministic mapping function $\mathsf{P}_e^{X}(\vec{v}_c)$ in (11.a) are certain. Further experimentation involving the *unknowable* random environmental ($\vec{V}_E$) and position ($\vec{V}_t$) estimation effects in sensing are best studied statistically. Thus, the respective estimated random input parameters of $\vec{V}_E$ and $\vec{V}_t$ are introduced resulting in the mapping to the random signature process $\vec{\mathbf{X}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$. The sensing uncertainty is then subsequently propagated into the random feature process $\vec{\mathbf{Y}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ and ultimately to the decision process $\mathbf{Q}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$. For brevity, $\vec{\mathbf{Y}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ is written as $\vec{\mathbf{Y}}$ and $\mathbf{Q}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ written as $\mathbf{Q}$.

The distributions associated with the input parameters in $\vec{V}_E$ and $\vec{V}_t$ are estimated from experimental data. The estimated parameters become factors within a Monte Carlo simulation.

The cumulative link information loss as quantified within (5) and defined in (9a), (9b), (9c), then become random variables as shown below.

$$\mathbf{IL}_{\vec{\mathbf{X}}} \approx \mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H};\vec{\mathbf{X}}\,(\vec{v}_c\,,\,\vec{V}_E\,,\,\vec{V}_t\,)); \tag{17.a}$$

$$\mathbf{IL}_{\vec{\mathbf{Y}}} \approx \mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H};\vec{\mathbf{Y}}\,(\vec{v}_c\,,\,\vec{V}_E\,,\,\vec{V}_t\,)); \tag{17.b}$$

$$\mathbf{IL}_{\mathbf{Q}} \approx \mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H};\mathbf{Q}\,(\vec{v}_c\,,\,\vec{V}_E\,,\,\vec{V}_t\,)) \tag{17.c}$$

Similarly, the link information loss $\mathbf{IL}_{S\Delta}$, $\mathbf{IL}_{F\Delta}$, and $\mathbf{IL}_{D\Delta}$ in (10.a), (10.b) and (10.c) also become random variables.

The unknowable characteristics of the observed signature process $\vec{\mathbf{X}}$ are realized within the input variables to the modeled training process $\vec{\mathbf{X}}'(\vec{v}_c\,',\vec{V}_E\,',\vec{V}_t\,')$. If we assume that $\vec{v}_c\,' \neq \vec{v}_c$, $\vec{V}_E\,' \neq \vec{V}_E$, $\vec{V}_t\,' \neq \vec{V}_t$, then the mapping to the non-optimal decision rule will be $d(\vec{v}_c\,',\vec{V}_E\,',\vec{V}_t')$ which will be written as ***d*** for brevity. The decision rule ***d*** is applied to $\vec{\mathbf{Y}}(\vec{v}_c,\vec{V}_E,\vec{V}_t)$ generating $\mathbf{Q'}(\vec{v}_c\,',\vec{V}_E\,',\vec{V}_t')$, written as **Q'**, while the optimal decision rule $\boldsymbol{d_{opt}}$ generates $\mathbf{Q}(\vec{v}_c,\vec{V}_E,\vec{V}_t)$. Each realization of ***d*** and $\boldsymbol{d_{opt}}$ resulting from each ensemble $\vec{\mathbf{X}}'(\vec{v}_c\,',\vec{V}_E\,',\vec{V}_t\,')$ and $\vec{\mathbf{X}}(\vec{v}_c,\vec{V}_E,\vec{V}_t)$ respectively in the Monte-Carlo simulation will result in the randomization of the imperfect training loss function in equation (18)

$$\mathbf{IL}_{T\Delta} \equiv \mathrm{I}(\mathbf{H};\mathbf{Q}) - \mathbf{I}(\mathbf{H};\mathbf{Q'}) \tag{18}$$

and the randomization of the cumulative loss function in (19)

$$\mathbf{IL_{Q'}} \approx \mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H;Q'}) \tag{19}$$

In (19), the special case of $\vec{V}_E' = \vec{V}_E$ and $\vec{V}_t' = \vec{V}_t$, the loss due to the optimal training of $\boldsymbol{d} = \boldsymbol{d_{opt}}$ yields $\mathbf{IL_{T\Delta}} = 0$ and thus $\mathbf{IL_{Q'}} = \mathbf{IL_Q}$. To narrow the focus of analysis, the training space ($\vec{v}_c'$, $\vec{V}_E'$, $\vec{V}_t'$) will be considered fixed and thus will become a component of the system control parameter $\vec{v}_c$. Therefore $\boldsymbol{d}$ becomes fixed by design as d.

3.7.2.1 Independent Sources of Uncertainty Loss

Loss due to isolated sources of uncertainty within the channel can be computed to provide a means to characterize the relative impacts to information flow at various points in the channel. The various sources of sensing uncertainty induce information loss in the channel as characterized by the random link loss functions $\mathbf{IL_{S\Delta}}$, $\mathbf{IL_{F\Delta}}$, $\mathbf{IL_{D\Delta}}$, and $\mathbf{IL_{T\Delta}}$. The prior distributions on the random parameters within $\vec{V}_E$ and $\vec{V}_t$ are propagated to the respective loss functions using Monte Carlo simulation.

Definition III: The expected value of the link information loss can be written as the expected values of the individual random loss components as in (20.a) – (20.c).

$$\mu_{\mathbf{IL}_{S_\Delta}} = \mathrm{E}\ \{\ \mathbf{IL_{S\Delta}}\} \tag{20.a}$$

$$\mu_{\mathbf{IL}_{F_\Delta}} = \mathrm{E}\ \{\ \mathbf{IL_{F\Delta}}\} \tag{20.b}$$

$$\mu_{\mathbf{IL}_{D_\Delta}} = \mathrm{E}\ \{\ \mathbf{IL_{D\Delta}}\} \tag{20.c}$$

$$\mu_{\mathbf{IL}_{T_\Delta}} = \mathrm{E}\ \{\ \mathbf{IL_{T\Delta}}\} \tag{20.d}$$

The mean total channel loss given (21) follows from the linearity of the expectation operation and the additive relationship between the formerly deterministic quantities in (20.a-20.d).

$$\mu_{\mathbf{IL}_{\mathrm{Total}}} = \mu_{\mathbf{IL}_{S_\Delta}} + \mu_{\mathbf{IL}_{F_\Delta}} + \mu_{\mathbf{IL}_{D_\Delta}} + \mu_{\mathbf{IL}_{T_\Delta}} \;. \tag{21}$$

The sensing uncertainty factors within $\vec{V}_E$ and $\vec{V}_t$ are assumed to be independent. Given that the total loss function $\mathbf{IL}_{\mathrm{Total}}$ can account for multiple independent sources of uncertainty within the parameter space of $(\vec{v}_c, \vec{V}_E, \vec{V}_t)$, the variance on $\mathbf{IL}_{\mathrm{Total}}$ is the sum of the individual variances within the components of $\mathbf{IL}_{\mathrm{Total}}$.

Corollary II: Assuming n_e factors within $\vec{V}_E$ and n_t factors within $\vec{V}_t$, the link loss variance can be decomposed as given in (22a), (22b), (22c) and (22d).

$$\sigma^2_{\mathbf{IL}_{S_\Delta}} = \sigma^2_{\mathbf{IL}_{S_\Delta(V_{E1})}} + \ldots \sigma^2_{\mathbf{IL}_{S_\Delta(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{S_\Delta(V_{t1})}} + \;\ldots \sigma^2_{\mathbf{IL}_{S_\Delta(V_{n_t})}} \tag{22.a}$$

$$\sigma^2_{\mathbf{IL}_{F_\Delta}} = \sigma^2_{\mathbf{IL}_{F_\Delta(V_{E1})}} + \ldots \sigma^2_{\mathbf{IL}_{F_\Delta(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{F_\Delta(V_{t1})}} + \ldots \sigma^2_{\mathbf{IL}_{F_\Delta(V_{n_t})}} \tag{22.b}$$

$$\sigma^2_{\mathbf{IL}_{D_\Delta}} = \sigma^2_{\mathbf{IL}_{D_\Delta(V_{E1})}} + \ldots \sigma^2_{\mathbf{IL}_{D_\Delta(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{D_\Delta(V_{t1})}} + \ldots \sigma^2_{\mathbf{IL}_{D_\Delta(V_{n_t})}} \tag{22.c}$$

$$\sigma^2_{\mathbf{IL}_{T_\Delta}} = \sigma^2_{\mathbf{IL}_{T_\Delta(V_{E1})}} + \ldots \sigma^2_{\mathbf{IL}_{T_\Delta(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{T_\Delta(V_{t1})}} + \ldots \sigma^2_{\mathbf{IL}_{T_\Delta(V_{n_t})}} \tag{22.d}$$

Definition IV: The expected value of the *cumulative* link information loss can then be written as the expected values of the individual random cumulative loss components as in (23.a) – (23.c).

$$\mu_{\mathbf{IL}_{\bar{x}}} = \mathrm{E}\,\{\, \mathbf{IL}_{\bar{\mathbf{x}}} \,\} \tag{23.a}$$

$$\mu_{\mathbf{IL}_{\vec{Y}}} = \mathrm{E}\{\,\mathbf{IL}_{\vec{\mathbf{Y}}}\,\} \tag{23.b}$$

$$\mu_{\mathbf{IL}_{\mathbf{Q}}} = \mathrm{E}\{\,\mathbf{IL}_{\mathbf{Q}}\,\} \tag{23.c}$$

$$\mu_{\mathbf{IL}_{\mathbf{Q}'}} = \mathrm{E}\{\,\mathbf{IL}_{\mathbf{Q}'}\,\} \tag{23.d}$$

Corollary III: Assuming $\mathsf{n_e}$ factors within $\vec{V}_E$ and $\mathsf{n_t}$ factors within $\vec{V}_t$, the *cumulative* link loss variance can be decomposed as given in (24.a), (24.b), (24.c) and (24.d).

$$\sigma^2_{\mathbf{IL}_{\vec{X}}} = \sigma^2_{\mathbf{IL}_{\vec{X}(V_{E1})}} + \dots \sigma^2_{\mathbf{IL}_{\vec{X}(V_{n_e})}} + \sigma^2_{\mathbf{IL}\vec{X}(V_{t1})} + \dots \sigma^2_{\mathbf{IL}_{\vec{X}(V_{n_t})}} \tag{24.a}$$

$$\sigma^2_{\mathbf{IL}_{\vec{Y}_\Delta}} = \sigma^2_{\mathbf{IL}_{\vec{Y}(V_{E1})}} + \dots \sigma^2_{\mathbf{IL}_{\vec{Y}(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{\vec{Y}(V_{t1})}} + \dots \sigma^2_{\mathbf{IL}_{\vec{Y}(V_{n_t})}} \tag{24.b}$$

$$\sigma^2_{\mathbf{IL}_{\mathbf{Q}}} = \sigma^2_{\mathbf{IL}_{\mathbf{Q}(V_{E1})}} + \dots \sigma^2_{\mathbf{IL}_{\mathbf{Q}(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{\mathbf{Q}(V_{t1})}} + \dots \sigma^2_{\mathbf{IL}_{\mathbf{Q}(V_{n_t})}} \tag{24.c}$$

$$\sigma^2_{\mathbf{IL}_{\mathbf{Q}'}} = \sigma^2_{\mathbf{IL}_{\mathbf{Q}'(V_{E1})}} + \dots \sigma^2_{\mathbf{IL}_{\mathbf{Q}'(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{\mathbf{Q}'(V_{t1})}} + \dots \sigma^2_{\mathbf{IL}_{\mathbf{Q}'(V_{n_t})}} \tag{24.d}$$

3.7.2.2 Propagating Link Loss to Link Performance

The variance and mean of the random cumulative loss components $\mathbf{IL}_{\vec{\mathbf{X}}}$, $\mathbf{IL}_{\vec{\mathbf{Y}}}$, $\mathbf{IL}_{\mathbf{Q}}$ and $\mathbf{IL}_{\mathbf{Q}'}$ are used directly to determine the variance on the performance at the random link performance components $\mathbf{P}_{\mathbf{e}}^{\vec{X}}$, $\mathbf{P}_{\mathbf{e}}^{\vec{Y}}$, $\mathbf{P}_{\mathbf{e}}^{\mathbf{Q}}$, and $\mathbf{P}_{\mathbf{e}}^{\mathbf{Q}'}$. The Maximum Likelihood Estimate (MLE) of $\mathsf{P_e}$ is inferred at each realization of the sufficient sample support about $(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ providing the random mapping to performance $\mathsf{P_e}$ at each link.

Corollary IV: Given sufficient sampling of the space of $\vec{V}_E$ and $\vec{V}_t$ within the finite alphabet $|\chi|$ and $|Y|$, the environmental and position estimate uncertainty factors result in the respective random performance at $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ given by functions $\mathrm{P_e}^{X}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ and $\mathrm{P_e}^{Y}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ as in equations (25) and (26).

$$\mathbf{P}_{\mathbf{e}}^{\vec{\mathbf{X}}} \equiv \mathrm{P_e}^{\mathrm{X}} \left(\vec{v}_c, \vec{V}_E, \vec{V}_t \right) \approx \mathrm{F}(\mathbf{IL}_{\vec{\mathrm{X}}}) \tag{25}$$

$$\mathbf{P}_{\mathbf{e}}^{\vec{\mathbf{Y}}} \equiv \mathrm{P_e}^{\mathrm{Y}} \left(\vec{v}_c, \vec{V}_E, \vec{V}_t \right) \approx \mathrm{F}(\mathbf{IL}_{\vec{\mathbf{Y}}}) \tag{26}$$

If conditions of Corollary IV hold and perfect training conditions are assumed where $\vec{v}_c' = \vec{v}_c$, $\vec{V}_E' = \vec{V}_E$, $\vec{V}_t' = \vec{V}_t$, then the mapping to the decision rule d_{opt} will be optimal.

Corollary V: The output of the discrete random variable $\mathbf{Q}$ (from the finite alphabet $|Q|$) is driven by the inferred decision out of the application of each realization of $\vec{\mathbf{Y}}$ to d_{opt}. The random performance function $\mathrm{P_e}^{\mathbf{Q}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ can be expressed as random realization of the information loss in the channel, $\mathbf{IL_Q}$ in (17c). Using the approximation form of (13) (assume $\mathrm{I}(\mathbf{Q;V}) \approx 0$), the random performance function $\mathbf{P_e}^{\mathbf{Q}}$ is given by (27).

$$\mathbf{P_e}^{\mathbf{Q}} \equiv \mathrm{P_e}^{\mathrm{Q}}(\vec{v}_c, \vec{V}_E, \vec{V}_t) \approx \mathrm{F}\{\mathbf{IL_Q}\} \tag{27}$$

The approximation in (27) can be replaced by an equality using the full representation in (4).

$$\mathbf{P_e}^{\mathbf{Q}} = \mathrm{F}\{\ \mathbf{IL_Q} + \mathrm{I}(\mathbf{Q};\mathbf{V})\} \tag{28}$$

In (27) and (28), the relaxation of the constraint $\vec{V}_E{}' = \vec{V}_E$ and $\vec{V}_t{}' = \vec{V}_t$ expands the study of the effects of uncertainty to the loss due to the non-optimal training of *d*.

Corollary VI: The output of the discrete random variable **Q'** (from the finite alphabet $|Q'|$) is driven by the inferred decision out of the application of each realization of $\vec{\mathbf{Y}}$ to *d*. The random performance function $\mathrm{P_e}^{\mathbf{Q'}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ can be expressed as random realization of the information loss in the channel, H(**H**) - I(**H;Q'**). Fixing the suboptimal decision rule *d*($\vec{v}_c{}' = \beta_c, \vec{V}_E{}' = \beta_E, \vec{V}_t{}' = \beta_t$) and using the form of (4) and assuming I(**Q;V'**) ≈ 0, the random performance function $\mathbf{P_e}^{\mathbf{Q'}}$ is given by (29).

$$\mathbf{P_e}^{\mathbf{Q'}} \equiv \mathrm{P_e}^{\mathrm{Q'}}(\vec{v}_c, \vec{V}_E, \vec{V}_t) \approx \mathrm{F}\{\mathbf{IL}_{\mathbf{Q'}}\} = \mathrm{F}\{\mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H;Q'})\} \tag{29}$$

The approximation in (29) is replaced by an equality using the full representation in (4).

$$\mathbf{P_e}^{\mathbf{Q'}} \equiv \mathrm{P_e}^{\mathrm{Q'}}(\vec{v}_c, \vec{V}_E, \vec{V}_t) = \mathrm{F}\{\mathrm{H}(\mathbf{H}) - \mathrm{I}(\mathbf{H;Q'}) + \mathrm{I}(\mathbf{Q'};\mathbf{V'})\} \tag{30}$$

Definition V: The expected link performance under control parameters $\vec{v}_c$ and in the presence of sensing uncertainty $(\vec{V}_E, \vec{V}_t)$ is defined as the expectation of the random link performance components $\mathbf{P_e^{\vec{X}}}, \mathbf{P_e^{\vec{Y}}}, \mathbf{P_e^{Q}}$, and $\mathbf{P_e^{Q'}}$.

$$\mu_{\mathbf{P_e}_{\vec{X}}} = \mathrm{E}\,\{\mathbf{P_e^{\vec{X}}}\} \tag{31.a}$$

$$\mu_{\mathbf{P_e}_{\vec{Y}}} = \mathrm{E}\,\{\mathbf{P_e^{\vec{Y}}}\} \tag{31.b}$$

$$\mu_{\mathbf{P_e}_{\mathbf{Q}}} = \mathrm{E}\,\{\mathbf{P_e^{Q}}\} \tag{31.c}$$

$$\mu_{\mathbf{P_e}_{\mathbf{Q'}}} = \mathrm{E}\,\{\mathbf{P_e^{Q'}}\} \tag{31.d}$$

Given a sufficient number of Monte-Carlo samples over the random parameters in $\vec{V}_E$ and $\vec{V}_t$, the standard deviation of the random link component performance function is used as a measure of reliability. Reliability is interpreted as the confidence that a classification event would result in performance that would fall within the bounds of one standard deviation of the mean performance.

> **Definition VI:** *Reliability* in predicted link performance is defined as the standard deviation ($\sigma_{\mathsf{P}_e^{\vec{X}}}$, $\sigma_{\mathsf{P}_e^{\vec{Y}}}$, $\sigma_{\mathsf{P}_e^{Q}}$, and $\sigma_{\mathsf{P}_e^{Q'}}$) of the respective random cumulative link performance associated with $\mathbf{P}_e^{\vec{X}}, \mathbf{P}_e^{\vec{Y}}, \mathbf{P}_e^{Q}$, and $\mathbf{P}_e^{Q'}$. The *variability* in link performance is defined as the square of the reliability; $\sigma^2_{\mathsf{P}_e^{\vec{X}}}$, $\sigma^2_{\mathsf{P}_e^{\vec{Y}}}$, $\sigma^2_{\mathsf{P}_e^{Q}}$, and $\sigma^2_{\mathsf{P}_e^{Q'}}$.

3.8 Uncertainty in Performance

The independent sources of uncertainty contributing to $\sigma^2_{\mathsf{IL}_{\vec{X}}}$ in (24.a) are individually functionally mapped to the variance on the random performance function $\mathbf{P}_e^{\vec{X}}$ to determine the respective effects on the reliability of the predicted link performance estimate. The uncertainty is passing through the transcendental relationship between $\mathsf{IL}_{\vec{X}}$ and $\mathsf{P}_e^{\vec{X}}$. The nature of the nonlinear relationship makes it difficult to commute the independent loss variance sources analytically. It is important to relate the independent sources of uncertainty underlying $\sigma^2_{\mathsf{IL}_{\vec{X}}}$ to the corresponding set of variances that combine to equal the variance on $\mathbf{P}_e^{\vec{X}}$.

While this relationship is transcendental and nonlinear, when the uncertainty is small and tight about the mean, it is possible to approximate the inverse entropy function (F) by a linear relationship [25] about the mean of $\mathbf{IL}_{\vec{X}}$ [4].

$$F(\mathbf{IL}_{\vec{x}}) = a + b\cdot(\mathbf{IL}_{\vec{x}})$$

The mean and variance of the of the approximation are then

$$E[F(\mathbf{IL}_{\vec{x}})] = a + b\cdot(\mu_{\mathbf{IL}_{\vec{x}}})$$

$$Var[F(\mathbf{IL}_{\vec{x}})] = b^2\cdot(\sigma^2_{\mathbf{IL}_{\vec{x}}})$$

Using established approximation techniques, the first order Taylor expansion of F around the mean $\mu_{\mathbf{IL}_{\vec{x}}}$ of $\mathbf{IL}_{\vec{x}}$ is equal to

$$F(\mathbf{IL}_{\vec{x}}) \approx F(\mu_{\mathbf{IL}_{\vec{x}}}) + F'(\mu_{\mathbf{IL}_{\vec{x}}})\cdot(\mathbf{IL}_{\vec{x}} - \mu_{\mathbf{IL}_{\vec{x}}}). \tag{32}$$

Using the Taylor Series expansion in (32), the approximations for $E[F(\mathbf{IL}_{\vec{x}})]$ and $Var[F(\mathbf{IL}_{\vec{x}})]$ are [4]

$$E[F(\mathbf{IL}_{\vec{x}})] = E[\mathbf{P}_e^{\vec{x}}] \approx F(\mu_{\mathbf{IL}_{\vec{x}}}) = H^{-1}(\mu_{\mathbf{IL}_{\vec{x}}}). \tag{33}$$

$$Var[F(\mathbf{IL}_{\vec{x}})] = \sigma^2_{\mathbf{P}_e^{\vec{x}}} \approx \{F'(\mu_{\mathbf{IL}_{\vec{x}}})\}^2\cdot(\sigma^2_{\mathbf{IL}_{\vec{x}}}) \tag{34}$$

and $F'(\mu_{\mathbf{IL}_{\vec{x}}})$ can be shown to equal

$$F'(\mu_{\mathbf{IL}_{\vec{x}}}) = \log\left[\frac{1}{\left[\frac{H^{-1}(\mu_{\mathbf{IL}_{\vec{x}}})}{1-H^{-1}(\mu_{\mathbf{IL}_{\vec{x}}})}\right]}\right].$$

Assuming n_e factors within $\vec{V}_E$ and n_t factors within $\vec{V}_t$, the *cumulative* link loss variance components given in (24.a) are applied to (34).

$$\sigma^2_{\mathbf{P}_e^{\vec{x}}} \approx \{F'(\mu_{\mathbf{IL}_{\vec{x}}})\}^2.\left\{\sigma^2_{\mathbf{IL}_{\vec{x}(V_{E1})}} + \ldots\sigma^2_{\mathbf{IL}_{\vec{x}(V_{n_e})}} + \sigma^2_{\mathbf{IL}_{\vec{x}(V_{t1})}} + \ldots\sigma^2_{\mathbf{IL}_{\vec{x}(V_{n_t})}}\right\} \tag{35}$$

The variance on the performance estimate $\mathbf{P}_e^{\vec{\mathbf{X}}}$ is then decomposed into the individual sources of sensing uncertainty being propagated through the decision space at $\vec{\mathbf{X}}$.

$$\sigma^2_{\mathbf{P}_e^{\vec{\mathbf{X}}}} \approx \sigma^2_{\mathbf{P}_{e_E}^{\vec{\mathbf{X}}}} + \ldots \sigma^2_{\mathbf{P}_{e_{n_e}}^{\vec{\mathbf{X}}}} + \sigma^2_{\mathbf{P}_{e_t}^{\vec{\mathbf{X}}}} + \ldots \sigma^2_{\mathbf{P}_{e_{n_t}}^{\vec{\mathbf{X}}}} \tag{36}$$

Similar methods are applied to the independent contributions to the sensing uncertainty of $\vec{V}_E$ and $\vec{V}_t$ comprising the variances $\sigma^2_{\mathbb{L}_{\vec{\mathbf{Y}}_\Delta}}$, $\sigma^2_{\mathbb{L}_{\mathbf{Q}}}$, and $\sigma^2_{\mathbb{L}_{\mathbf{Q}'}}$ at $\vec{\mathbf{Y}}$, **Q**, and **Q'** respectively.

3.8.1 Stability of the Linear Approximation

The validity of the linear approximation in (34) requires $\sigma^2_{\mathbb{L}_{\vec{\mathbf{X}}}}$ be small. Thus, the contributing sources of sensing uncertainty within $\sigma^2_{\mathbb{L}_{\vec{\mathbf{X}}}}$ must be individually small. Given that the regime of interest is one where $\mu_{\mathbb{L}_{\vec{\mathbf{X}}}}$ and thus E[$\mathbf{P}_e^{\vec{\mathbf{X}}}$] are small, the derivative (slope) evaluated at $\mu_{\mathbb{L}_{\vec{\mathbf{X}}}}$ is relatively small. The slope within this regime is illustrated in Fig. 6 for an arbitrary operating point.

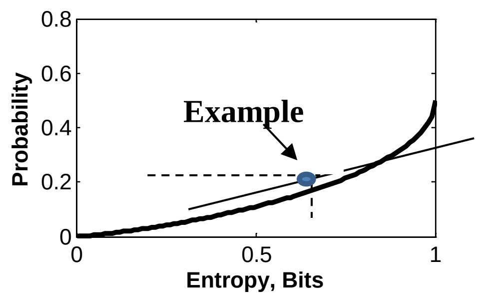


Fig. 6. The inverse Entropy Function; $f(z) = \mathrm{H}^{-1}(w)$

The slope $\frac{d\,\mathrm{H}^{-1}(w)}{dw}$ is plotted in Fig. 7 for $w \in [0, 1]$.

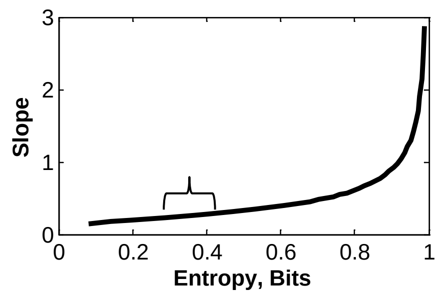


Fig. 7. The Derivative, $\frac{d\,\mathrm{H}^{-1}(w)}{dw}$, versus H($z$)

ATR design solutions of interest are typically in the range where $\mathrm{P_e} < .1$. From Fig. 7 it is evident that the slope at an operating point within this regime will be in the range [0,.25] of 0.25 affording reduced sensitivity to effects of the size of $\sigma^2_{\mathbb{L}_{\vec{x}}}$.

3.9 Dimensionality and Computing

The computation of the entropy of $\vec{\mathrm{x}}$ involves the joint probability mass function (PMF) of the random multivariate $\vec{\mathrm{x}}$ and is complicated by the large dimensional nature of the observation mapping $\mathrm{H} \rightarrow \vec{\mathrm{x}}$. It is desired to compute the discrete entropy for $\vec{\mathrm{x}}$ absent any assumption regarding dependence between the respective dimensions of $\vec{\mathrm{x}}$. If the $\vec{\mathrm{x}}$ space consists of K random variables or indices (dependent or independent) and the random variable $\mathrm{x}_k; k \in \{1,\mathrm{K}\}$ has $\mathrm{n_b}$ distinct bins (statistical divisions), then the size of the alphabet of $\vec{\mathrm{x}}, |\vec{\chi}|$, is given in (37) below.

$$|\vec{\chi}| = \prod_{k=1}^{k=K} \mathrm{n_k} \,. \tag{37}$$

For example, if $\kappa = 3$ and $\mathrm{n}_k = 2 = \mathrm{n}_b$ for all k, $|\vec{\chi}| = 2 \bullet 2 \bullet 2 = 8$.

The joint PMF of $\vec{\mathsf{x}}$, $p(x^j_{k_N})$; $k \in \{1, \mathsf{K}\}, j \in \{1, \mathsf{n_b}\}$ is generated from a finite N sample ensemble and discretely binned with $\mathsf{n_b}$ statistical divisions within each of the K elements of $\vec{\mathsf{x}}$. Stable entropic estimates require the statistics of the multivariate PMF be sampled sufficiently. A reasonable example in the context of the HRR example with K=10 and $\mathsf{n_b}$ =5 for all *k* would present a theoretical upper bound on the typical set of 5^{10}=9,765,625 [2]. The typical set represents the set of most probable events and contains almost all of the probability as the number of samples increases. In the case of the radar example developed here, this would be the set of most probable signature amplitude combinations for all K dimensions of $\vec{\mathsf{x}}$. To generate a meaningful sample size for a PMF of this size, we would need to produce at least 10 times the actual typical set. This means we need approximately *100 million samples.* Thus κ and $\mathsf{n_b}$ drive the dimensionality of $\vec{\mathsf{x}}$ and subsequently the sampling requirements for each ensemble within the Monte-Carlo simulation.

A high dimensional problem is one where the alphabet of $\vec{\mathsf{x}}$, $|\vec{\chi}|$, underlying the random process far exceeds the number of samples observed (N), i.e.; $|\vec{\chi}| >> \mathsf{N}$. Sensing systems typically operate within this high dimensional signature data space of $|\vec{\chi}|$. The high dimension arises due to factors within the space $\vec{\mathsf{x}}(\vec{v}_c, \vec{V}_E, \vec{V}_t)$. Hypothesis testing and inference within the high dimensional space of $\vec{\mathsf{x}}$ in turn leads to large sampling requirements to adequately determine the underlying statistical nature of the phenomenon under study. Without accurate determination of the underlying system statistics, poorly performing hypothesis tests and/or parameter estimation occur (Bias/Variance tradeoff) [34].

The number of statistical bins, $\mathsf{n_b}$, within the discrete sampling of the κ element joint PMF of $\vec{\mathsf{x}}$ also has a significant effect on $|\vec{\chi}|$ and thus on the entropy computation of $\vec{\mathsf{x}}$. An increase in size of $\mathsf{n_b}$ in $\vec{\mathsf{x}}$ will result in an increase in the entropy of $\vec{\mathsf{x}}$. However, in the limit, the value for I($\mathbf{H}$; $\vec{\mathsf{x}}$) as a function of $\mathsf{n_b}$ asymptotes to a constant value - after one reaches the full intrinsic dimensionality of the subspace of I($\mathbf{H}$; $\vec{\mathsf{x}}$) [35]. This will be true for (I($\mathbf{H}$; $\vec{\mathsf{Y}}$), I($\mathbf{H}$;$\mathbf{Q}$), and I($\mathbf{H}$;$\mathbf{Q'}$) as well. Choosing the most challenging link in the channel and without loss of generality, a method for determining the intrinsic dimensionality of $\vec{\mathbf{X}}$ is then needed to guide the selection of N.

3.9.1 Sample Size

The subject of nonparametric estimation of discrete entropy and mutual information of random variables have been widely published [36]. The intent of this section is not to expand this body of knowledge but to record the approach employed to determine the minimum sampling requirements for the entropy estimation. The variance parameters of these estimates is of particular interest.

The link performance variability estimate at each of the respective links, $\sigma^2_{\mathrm{P}_e^{\vec{X}}}, \sigma^2_{\mathrm{P}_e^{\vec{Y}}}, \sigma^2_{\mathrm{P}_e^{Q}}$, and $\sigma^2_{\mathrm{P}_e^{Q'}}$ are generated through a sufficient number of draws from the respective random link performance functions $\mathbf{P}_e^{\vec{X}}, \mathbf{P}_e^{\vec{Y}}, \mathbf{P}_e^{Q}$, and $\mathbf{P}_e^{Q'}$. Each draw involves the estimation of an entropic quantity computed from PMF $p(x_{k_N}^j)$ based on the N sample ensemble taken from $\vec{\mathbf{X}}$. The estimate of the link performance variability at $\vec{\mathbf{X}}$, $\hat{\sigma}^2_{\mathrm{P}_e^{\vec{X}}}$, is written more precisely as in (38) below.

$$\hat{\sigma}^2_{\mathrm{P}_e^{\vec{X}}} = \sigma^2_{\mathrm{P}_e^{\vec{X}}} + \sigma^2_{\hat{\mathrm{P}}_{e_N}^{\vec{X}}} \tag{38}$$

$\sigma^2_{\hat{\mathrm{P}}_{e_N}^{\vec{X}}}$ is defined as the N sample estimation variance or "sampling uncertainty" associated with the true variability $\sigma^2_{\mathrm{P}_e^{\vec{X}}}$. Equation (38) can be written as

$$\hat{\sigma}^2_{\mathrm{P}_e^{\vec{X}}} = \sigma^2_{\mathrm{P}_e^{\vec{X}}}\left(1 + \sigma^2_{\hat{\mathrm{P}}_{e_N}^{\vec{X}}} \Big/ \sigma^2_{\mathrm{P}_e^{\vec{X}}}\right).$$

For the high dimensional problem, N must be large enough for

$$\frac{\sigma^2_{\hat{\mathbf{P}}^{\vec{\mathbf{X}}}_{e_{\mathrm{N}}}}}{\sigma^2_{\mathbf{P}^{\vec{\mathbf{X}}}_{e}}} << 1 \cdot \qquad 39)$$

The objective then is to produce link reliability estimates that are within this regime. The choice of N must be selected to ensure the uncertainty of the entropic estimate is much less than the reliability limits realized due to various factors within $(\vec{v}_c, \vec{V}_E, \vec{V}_t)$ under study. That is, the ensemble size N of $\vec{\mathbf{X}}$, $\vec{\mathbf{Y}}$, $\boldsymbol{Q}$, and $\boldsymbol{Q}$' should be sufficiently large to ensure that the variance of the estimate falls within three significant digits of the variability levels $(\sigma^2_{\mathbf{P}^{\vec{\mathbf{X}}}_{e}}, \sigma^2_{\mathbf{IL}_{\vec{\mathbf{Y}}_\Delta}}, \sigma^2_{\mathbf{IL}_{\mathbf{Q}}}$, and $\sigma^2_{\mathbf{IL}_{\mathbf{Q}'}}$). Thus for the case of variability at $\vec{\mathbf{X}}$ we desire $\frac{\sigma^2_{\hat{\mathbf{P}}^{\vec{\mathbf{X}}}_{e_{\mathrm{N}}}}}{\sigma^2_{\mathbf{P}^{\vec{\mathbf{X}}}_{e}}} < .001 \cdot$

As stated above, $|\vec{\chi}|$ in particular, can grow to large levels and as such the number of samples required will grow as well. Given that the sampling ensemble size N of $\vec{\mathbf{X}}$ is the defining case, the following analysis is focused on the process at $\vec{\mathbf{X}}$ and this minimum N determination will be imposed also on $\vec{\mathbf{Y}}$, and $\boldsymbol{Q}$,

3.9.2 Phase Transitions and the Typical Set

The entropy computation [36] requires the development of the joint mass function associated with the multi-variate $\vec{\mathbf{X}}$, $p(x_{\mathrm{k}}^{\mathrm{j}})$; $\mathrm{j} \in \{1{:}\mathrm{n_b}\}$, $\mathrm{k} \in \{1{:}\mathrm{K}\}$. The development of this mass function assumes no independence between the K indices of $\vec{\mathbf{X}}$ and is performed using a “linked list” approach to limit the memory requirements during computation. A doubly linked list implementation with a hash table search approach yields a computational complexity of $O(\mathrm{N})$ [4]. The Miller Madow estimate [37] provides a faster convergence over the MLE method [2] for finite sample estimates.

Maximum Likelihood Estimate of $H(\vec{\mathbf{X}}_{\mathrm{k}})$;

$$\hat{\mathrm{H}}_{MLE_N}(\mathbf{X}_k) = \sum_{j=1}^{n_b} - p(x_{k_N}^j)\log_2\{p(x_{k_N}^j)\} \cdot \tag{40}$$

Miller-Madow Estimate of $\mathrm{H}(\vec{\mathbf{x}}_k)$; (note: M_+= number of statistical bins for which $p(x_{k_N}^j) \neq 0$)

$$\hat{\mathrm{H}}_{MM_N}(\mathbf{X}_k) = \hat{\mathrm{H}}_{MLE_N}(\mathbf{X}_k) + \{1/(2N)\}\{M_+ - 1\} \tag{41}$$

The N sample estimates for $\hat{\mathrm{H}}_{MLE_N}(\mathbf{X}_k)$ and $\hat{\mathrm{H}}_{MM_N}(\mathbf{X}_k)$ are generated from the joint mass function, $p(x_k^j)$; $j \in \{1{:}n_b\}$, $k \in \{1{:}K\}$.

Phase transitions [38]-[42] within the growth trajectory of the estimated entropy with increasing N are useful in defining the alphabet size $|\vec{\chi}|$. The following illustration demonstrates the usefulness of this approach. The signature process under evaluation is constructed by design such that the actual entropy value is known. We model the multivariate random signature vector $\vec{\mathbf{x}}$ to be uniformly distributed (standard uniform {0,1}) with n_b=6 (all indices of $\vec{\mathbf{x}}$) and K=3. The theoretical maximum value of the entropy of $\vec{\mathbf{x}}$ is then $\log_2(n_b^K)$ or $\log_2(6^3)$=7.7549 Bits. In Fig. 8 we incrementally generate the estimate of the discrete entropy of $\vec{\mathbf{x}}$ for an increasing number of samples. We plot the typical set of $\vec{\mathbf{x}}$ for each increment. The typical set $\mathbf{A}_\varepsilon = \mathbf{2}^{\mathrm{H}(\vec{\mathbf{x}})}$ is computed from the discrete entropy $H(\vec{\mathbf{x}})$. Each of the estimated values for the typical set of $\vec{\mathbf{x}}$ asymptote at the maximum dimensionality of $\vec{\mathbf{x}}$ where the theoretical values of $\mathrm{H}(\vec{\mathbf{x}})$ = 7.7549 Bits and $\mathbf{A}_\varepsilon$ = 216.

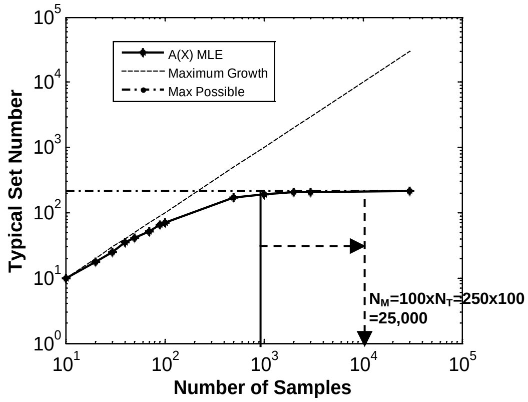


Fig. 8. Phase Transitions in $\vec{X}$ and Computing the Minimum Sampling N_M, MLE Method

Initially the samples are filling the open high dimensional space of $\vec{X}$ in a uniform fashion. The linear dashed line represents the $\log_2(N)$ growth of the entropy associated with this uniform distribution. Note that the actual achieved entropy computation begins to diverge from a uniform distribution. Only after the samples of $\vec{X}$ begin to accumulate in the bin space of the joint mass function of $\vec{X}$ does this transition occur. This phase transition point represents the point at which "collisions" occur and the fundamental statistics of $\vec{X}$ change.

The phase transition point is determined from intersection of the line tangent to the linear portion of the typical set profile and the line tangent to the asymptotic portion of the profile. The number of samples coinciding with this phase transition point is N_T. For the example here, N_T is found to be approximately 250 as illustrated in Fig. 8. The minimum number of samples, N_M, is taken to be 100 times the value of N_T. In this example N_M is found to be 25,000. The Miller-Madow estimate for entropy $\hat{H}_{MM_N}(X_k)$ is used for all entropic computation within the remaining body of this analysis.

3.10 Sampling Uncertainty for Probability of Error Estimate

Since the random estimation error variable is essentially the sum of many independently distributed random variables, the estimation error is Gaussian. The standard deviation of the Gaussian distribution of $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$, will then scale as a function of $1/N$. Thus the variance on the estimate $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$, $\sigma^2_{\hat{I}(H;x)_{2N_T}}$, can be scaled to large sample size ($\sigma^2_{\hat{I}(H;x)_N}$). The standard deviation of the estimate $\hat{P}_e^{\,x}$ can be determined from the independent contributions of $H(\mathbf{H})$ and $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$ shown in (42).

$$\hat{P}_e^{\,x} \approx H^{-1}(H(\mathbf{H}) - \hat{I}(\mathbf{H};\vec{\mathbf{X}})) \tag{42}$$

For the equal probable binary hypothesis case, $H(\mathbf{H})$ is equal to 1 Bit. Therefore the sampling uncertainty $\sigma^2_{\hat{P}^{\vec{x}}_{e_N}}$ is a function only of $\sigma^2_{\hat{I}(H;x)_N}$.

As in section III.H, the inverse entropy function in (42) is a transcendental function and as such the variance on the estimate $\hat{P}_e^{\,x}$, $\sigma^2_{\hat{P}^{\vec{x}}_{e_N}}$, can be very difficult to determine analytically. Following a similar line of analysis as in section III.H using (33) and (34), the mean and variance of $\hat{P}_e^{\,x}$ can be calculated as

$$E[\hat{P}_e^{\,x}] \approx H^{-1}(1-\mu_{\hat{I}(H;x)_{2N_T}}) \tag{43}$$

$$\sigma^2_{\hat{P}^{x}_{e_N}} \approx \{ f'(\mu_{\hat{I}(H;x)_{2N_T}})\}^2 \cdot \sigma^2_{\hat{I}(H;x)_N} . \tag{44}$$

The use of (44) requires an estimate of the mean of $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$ which is taken to be the sample mean $\mu_{\hat{I}(H;x)_N}$. The ultimate goal is to learn the sampling uncertainty for $\hat{P}_e^{\,x}$, $\sigma_{\hat{P}^{\vec{x}}_{e_N}}$, from a low sample estimate of the mean of $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$, $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$. Manipulating (44) above, $\sigma^2_{\hat{\mathrm{I}}(\mathrm{H;x})_N}$ can be written in terms of the required variance on the estimate of error, $\sigma^2_{\hat{P}_e^x{}_{Req}}$;

$$\sigma^2_{\hat{\mathrm{I}}(\mathrm{H;x})_N} \leq \sigma^2_{\hat{P}_e^x{}_{Req}} \cdot \log\left(\frac{H^{-1}(1-\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}})}{1-H^{-1}(1-\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}})}\right)^2 . \tag{45}$$

To ensure $\sigma_{\hat{P}e_N} \leq \sigma_{\hat{P}e_{Req}}$, the relationship in (45) is essential.

The regime of interest is where $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$ is close to 1 and $H^{-1}(1-\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}})$ and thus $\hat{P}_e{}^x$ is small. The derivative of the estimate in this regime is in the range of [0, .25] as illustrated in Fig. 7. A slope of less than .25 is small relative to the range of values given in Fig. 7 yet large with respect to $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$. Therefore, errors in the estimate of $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})}$ can have a significant impact on the estimate of the number of samples required to reach a target sampling uncertainty of $\sigma^2_{\hat{P}_e^x{}_{Req}}$. This means that a conservative approach is needed to estimate E[$\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$] based on a small number of samples. Instead of using the sample mean $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$ as an estimate of the expectation E{ $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$ }, a value somewhat less than the sample mean should be chosen. Depending on the level of confidence required in the estimate of the number of samples N, a higher confidence estimate can be achieved by replacing $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$ with $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}} - \sigma_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$ in (45).

As discussed above, the variance on the estimate $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$, $\sigma^2_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$, can be scaled to large sample size ($\sigma^2_{\hat{\mathrm{I}}(\mathrm{H;x})_N}$). The mean of the estimate of $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$, $\mu_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$, and the standard deviation, $\sigma_{\hat{\mathrm{I}}(\mathrm{H;x})_{2N_T}}$, can be estimated using the low number of samples ($N=2N_T$).

3.11 Sampling Uncertainty versus Variability in Performance

The expression in (45) provides guidance on the level of sampling uncertainty associated with $\hat{\mathrm{I}}(\mathbf{H};\vec{\mathbf{X}})$ required to achieve the corresponding sampling uncertainty in $\hat{\mathrm{P}}_{\mathrm{e}}^{\mathbf{X}}$. A more important question relevant to the study of uncertainty and performance estimation is the relationship introduced in (44) and written in general form below.

$$\frac{\sigma^2_{\hat{\mathrm{P}}^{\vec{\mathrm{X}}}_{e_{\mathrm{N}}}}}{\sigma^2_{\mathrm{P}^{\vec{\mathrm{X}}}_{e}}} < \alpha \tag{46}$$

The variable α can be set to limit the degree of sampling uncertainty to be realized in the performance confidence analysis. Using (44), previous development, and the fact that $\sigma^2_{\mathrm{IL}_{\vec{\mathrm{X}}}}$ = $\sigma^2_{\mathrm{I}(\mathbf{H};\vec{\mathbf{X}})}$; (46) can be written as in (47).

$$\frac{\sigma^2_{\hat{\mathrm{P}}^{\vec{\mathrm{X}}}_{e_{\mathrm{N}}}}}{\sigma^2_{\mathrm{P}^{\vec{\mathrm{X}}}_{e}}} \approx \left[\frac{\sigma^2_{\hat{\mathrm{I}}(\mathbf{H};\mathbf{X})_{2\cdot \mathrm{N_T}}}}{\sigma^2_{\mathrm{I}(\mathbf{H};\vec{\mathbf{X}})}}\right] \cdot \beta(\mathrm{N},\mathrm{N_T}) \tag{47}$$

The factor $\beta(\mathrm{N},\mathrm{N_T})$ in (47) is given as

$$\beta(\mathrm{N},\mathrm{N_T}) = \left[\frac{4\cdot \mathrm{N_T^2}}{\mathrm{N}^2}\right] \cdot \left[\frac{\log\left[\frac{\mathrm{H}^{-1}\left(1-\mu_{\mathrm{I}(\mathbf{H};\vec{\mathbf{X}})}\right)}{1-\mathrm{H}^{-1}\left(1-\mu_{\mathrm{I}(\mathbf{H};\vec{\mathbf{X}})}\right)}\right]}{\log\left[\frac{\mathrm{H}^{-1}\left(1-\mu_{\hat{\mathrm{I}}(\mathbf{H};\mathbf{X})_{2\cdot \mathrm{N_T}}}\right)}{1-\mathrm{H}^{-1}\left(1-\mu_{\hat{\mathrm{I}}(\mathbf{H};\mathbf{X})_{2\cdot \mathrm{N_T}}}\right)}\right]}\right]^{\cdot}$$

Thus the expression in (48) can be used to test for conditions specified in (46).

$$\left[\frac{\sigma^2_{\hat{\mathrm{I}}(\mathbf{H};\mathbf{X})_{2\cdot \mathrm{N_T}}}}{\sigma^2_{\mathrm{I}(\mathbf{H};\vec{\mathbf{X}})}}\right] \cdot \beta(\mathrm{N},\mathrm{N_T}) < \alpha \tag{48}$$

The FBIT model provides a platform for the study and analysis of the relationship of the level of sampling uncertainty to the level of performance uncertainty. Incremental values for the ratio on the left side of (48) can be computed for increasing N. In section IV, the point at which the inequality is obeyed is related to the phase transition minimum sample methods generated in section III.I.

4 AN INFORMATION FLOW NUMERICAL EXAMPLE

The application of the FBIT method to the study of uncertainty propagation is now illustrated within a simple radar sensor example. An information loss budget is constructed for a baseline design. Selected forms of uncertainty in Table I are introduced into the system to demonstrate the analysis of the effects of propagating uncertainty through the information sensing channel.

4.2 Observed Target Scattering Model

In the high frequency regime used to obtain HRR signatures, the target may be approximated as a collection of scattering centers valid over a limited aspect window and frequency band. These scattering centers may be considered to be localized to a point and may represent a variety of scattering phenomena ranging from specular reflection to diffraction phenomena such as edge and tip diffraction. The fields radiated by these point scatterers depend upon both temporal and spatial frequencies (angular dependence). Since the radar illuminating the target has finite bandwidth and is a one dimensional imaging system, the target is seen as a collection of contiguous swaths of range, with each range swath corresponding to a particular range. The extent of each range swath (range resolution) depends upon the signal bandwidth. For a typical extended target of interest, each range swath contains a number of scattering centers which can be widely spaced in cross-range [17].

The electromagnetic field obtained as a result of the interference of the scattered fields from the scattering centers appears as the signal corresponding to a particular range bin of the target signature. The target signature may be considered to be a one dimensional image of the reflectivity (or scattering) profile of the target for a given azimuth/elevation aspect angle (θ,ϕ) and bandwidth. The mathematical definition of the radar signature is developed from the normalized scattered field in (49). $\vec{E}^s$ and $\vec{E}^i$ are the scattered field and the incident field respectively.

$$S(\theta,\phi) = \lim_{R\to\infty} 4\pi R^2 \frac{\vec{E}^s}{\left|\vec{E}^i\right|} \tag{49}$$

Using scattering center modeling and the far field approximation, (49) can be written in terms of the target aspect angle and the transmitted wavelength as shown in (50) [43].

$$\mathsf{S}_{\mathsf{E}}(\theta,\phi,\lambda) = \sum_{m=1}^{M} \sqrt{\sigma_m}\; e^{j\frac{4\pi}{\lambda}R_m(\theta,\phi)} \tag{50}$$

In equation (50) S_{E} is the band-limited frequency response of the target comprised of M scattering centers at the respective range R_m. Conditioned on the target hypothesis **H** at a fixed aspect angle (θ_i,ϕ_i), $\vec{\mathsf{S}}_{\mathsf{E}}(\theta_i,\phi_i) = \mathsf{S}_{\mathsf{E}}(\theta_i,\phi_i,\lambda)\,;\lambda\in\{\lambda_l,\lambda_{l+1},\ldots\lambda_f\}$ defines the band-limited frequency response of the normalized scattered field measurements given in (50). Clusters of simple scattering centers are chosen for targets of interest at X-band frequencies (8-12 GHz) in the following development. The targets are electrically large with dimensions in range and cross-range of many wavelengths.

The target cluster of M isotropic scatters occupies the target volume within the coordinate system illustrated in Fig. 9.

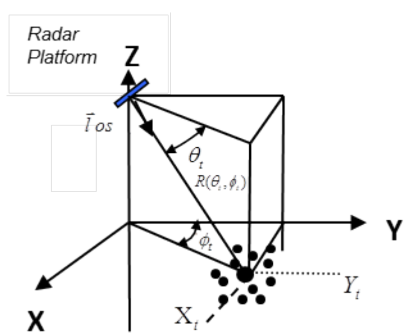


Fig. 9. Radar Sensor Coordinate System

The 3 dimensional target scattering center configuration for the two targets examined in the following example occupy an approximate cubic volume of {x=2, y=3, z=2.5} meters and are positioned at a line-of-site, $\vec{l}os$,of for $(\theta_t, \phi_t) = 10^{\circ}, 7.5^{\circ}$. Both targets are comprised of 100 scattering centers of unity amplitude and three strong localized scattering clusters of amplitude 5. Target 1 differs from target 2 in that the length of the target 1 is shorter than target 2 in the Y dimension by .5 meters. One of the localized scattering clusters is also displaced by (.2, .2, 0) meters.

4.3 Radar Sensor Model

Applying matched filter processing and the discrete Fourier transform to the observed signature $\vec{S}_E(\theta_i, \phi_i)$ in additive noise, the measured HRR signature can be modeled for a range of frequencies present in the transmitted waveform. The multidimensional encoded source $\vec{X}_E^i$ is defined here as the vector form of the time delay transformation of the band-limited frequency response $\vec{S}_E(\theta_i, \phi_i)$. The measured random signature process $\vec{\mathbf{X}}_n^i$ is then defined as in equation (51) were $\vec{\mathbf{n}}$ is additive white noise [17].

$$\vec{\mathbf{X}}_n^i = \vec{X}_E^i + \vec{\mathbf{n}} \tag{51}$$

The process $\vec{\mathbf{X}}_n^i$ is modeled at the output of a radar step frequency measurement sensor system for the specified target aspect angle (θ_i, ϕ_i). The additive noise process $\vec{\mathbf{n}}$ is modeled as the sum of thermal white noise and quantization noise components. The quantization error component is thought of as a random process uncorrelated with both the signal and the thermal noise. The complete radar step frequency measurement model system parameters are summarized below in Table III.

TABLE III. SENSOR SUMMARY

Center frequency	9.6 GHz
Transmit Bandwidth	800 MHz
Number Bits in A/D Conversion	8 Bits
Number of Pulses Integrated	1024
Signal-to-Noise Ratio (time delay domain)	20dB (variable)

The sensing of $\vec{X}_n^i$ in a dynamic real world environment is subject to the uncertainties listed in area 1 of Table I leading to the random signature process $\vec{X}$ as outlined in section III.E and summarized in Table II. Given the dynamic nature of the phenomenon underlying these uncertainties, the statistics associated with the dimensions of $\vec{X}$ are often time varying. The target statistics are assumed to be stationary (constant with time), thus the sample signatures associated with this random vector correspond to a stationary random process. Given the short measurement times associated with radar measurements of the nature under study, this assumption is judged as appropriate.

4.3.1 Modeling Pose Angle Estimation Uncertainty

The observed object aspect angle estimate can be viewed as lying within a solid cone angle centered on the observed object aspect angle (θ_t, ϕ_t). The parameter $\sigma_{\mathbf{t}}$ is defined as the uncertainty associated with the sensor estimate of (θ_t, ϕ_t). The parameter σ_t and μ_t are elements of $\vec{V}_t$ and are the standard deviation and bias of the object aspect angle estimate respectively.

The variation in measured signature phenomenology due to the uncertainties in target aspect angle are generated in the signal model in (50) through the introduction of distributions on θ and ϕ. The parameters θ and ϕ are both modeled as Gaussian random variables each with variance σ_t^2 and mean $\mu_t + \theta_t$, $\mu_t + \phi_t$. The bias parameter μ_t is assumed to be unknown and is modeled uniformly distributed between the interval [-1, 1] degrees.

4.3.2 Modeling Leading Edge Position Estimation Uncertainty

The target leading edge location estimation will vary under real world sensing conditions. Thus the range alignment (along the $\vec{l}os$) of the measured signature process $\vec{X}$ to the decision rule training process $\vec{X}'$ is imperfect and can be modeled as an uncertainty source. The process $\vec{X}$ alignment to $\vec{X}'$ is modeled through a positive bias applied to the phase center of the scattering cluster underlying $\vec{\mathbf{X}}$. The bias parameter μ_r is assumed to be unknown and is modeled uniformly distributed between [0, .2] meters. Note that μ_r is another element of $\vec{V}_t$.

4.3.3 Modeling Imperfect Training

The training process component $\vec{\mathbf{X}}'$ in Fig. 2 represents the best achievable statistical characterization of the observed signature process $\vec{\mathbf{X}}$. Signature training processes must represent the radar measured signature process across a wide range of measurement uncertainties and target configurations as well as under many uncertain operating conditions including clutter, obscuration, and other sources of RF interference. Construction of a signature training database derived entirely from measurements is expensive and can be an impractical proposition. It is possible to construct a signature database using electromagnetic scattering codes. However, given the complexity of typical targets and the challenge of modeling a variety of electromagnetic scattering phenomena ranging from specular reflection to edge diffraction, smooth surface diffraction etc., computation of signatures with sufficient accuracy is also a challenging task [17]. Within this analysis the dissimilarity between $\vec{\mathbf{X}}$ with $\vec{\mathbf{X}}'$ will be generated using a matched scattering center model configuration with $\vec{\mathbf{X}}$. The uncertain parameters of $\vec{V}_t$ and $\vec{V}_E$ modeled within $\vec{\mathbf{X}}$ are not modeled in $\vec{\mathbf{X}}'$. $\vec{\mathbf{X}}' = \vec{\mathbf{X}}$ only when $\vec{\mathbf{X}}$ is used directly for the training of the decision rule *d*.

4.4 Feature Discrimant and Decision Rule Design

The function *f* used to compute the feature discriminate $\vec{\mathbf{Y}}$ from $\vec{\mathbf{X}}$ in Fig. 2 is developed from the squared error of the distance from the mean templates $\vec{\mu}_{\vec{\mathbf{X}}_1}$ and $\vec{\mu}_{\vec{\mathbf{X}}_2}$ derived from the marginal training processes $\vec{\mathbf{X}}'_1$ and $\vec{\mathbf{X}}'_2$ as defined below [33]. The operator $|\vec{\tau}|$ is defined as the element-wise magnitude of each complex element of the random vector $\vec{\tau}$.

$$\vec{\mu}_{\vec{\mathbf{X}}'_1} = \mathrm{E}\left\{\left|\vec{\mathbf{X}}'_1\right|\right\},\ \vec{\mu}_{\vec{\mathbf{X}}'_2} = \mathrm{E}\left\{\left|\vec{\mathbf{X}}'_2\right|\right\},\ \vec{\mu}_{\vec{\mathbf{X}}'_{12}} = \vec{\mu}_{\vec{\mathbf{X}}'_1} - \vec{\mu}_{\vec{\mathbf{X}}'_2}$$

$$\vec{\mathbf{Y}}_1 = \left[\left|\vec{\mathbf{X}}_1\right| - \vec{\mu}_{\vec{X}'_{12}}\right] \cdot \left[\left|\vec{\mathbf{X}}_1\right| - \vec{\mu}_{\vec{X}'_{12}}\right]^T$$

$$\vec{\mathbf{Y}}_2 = \left[\left|\vec{\mathbf{X}}_2\right| - \vec{\mu}_{\vec{X}'_{12}}\right] \cdot \left[\left|\vec{\mathbf{X}}_2\right| - \vec{\mu}_{\vec{X}'_{12}}\right]^T$$

$$\vec{\mathbf{Y}} = [\vec{\mathbf{Y}}_1, \vec{\mathbf{Y}}_2]$$

The Maximum Likelihood estimator is used to determine the optimal decision rule *d*.

$$d = \mathrm{E}\left\{\frac{\mathrm{E}\{\vec{\mathbf{Y}}_1\} - \mathrm{E}\{\vec{\mathbf{Y}}_2\}}{2}\right\}$$

Assuming equally likely priors on each of the binary hypotheses H_1 and H_2 in $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$, the samples ($\vec{Y}$) from $\vec{\mathbf{Y}}$ are applied to the decision rule d. $\vec{Y} < d$ are declared from H_1 (denoted Q_1) and $\vec{Y} > d$ are declared from H_2 (denoted Q_2). The in-class and out-of-class scoring system is given by the conditional probabilities within $\alpha, \beta, \gamma,$ and κ as provided below.

$$\alpha = p(\vec{X}_1) \cdot p\left(Q_1 / \vec{X}_1\right), \quad \beta = p(\vec{X}_1) \cdot p\left(Q_2 / \vec{X}_1\right)$$

$$\gamma = p(\vec{X}_2) \cdot p\left(Q_1 / \vec{X}_2\right), \quad \kappa = p(\vec{X}_2) \cdot p\left(Q_2 / \vec{X}_2\right)$$

The output of the decision algorithm **Q** as formed from the scoring system above can be summarized by the confusion matrix for the binary classifier given in Fig. 10 below.

Test Class/ Train Class	$\vec{\mathbf{X}}'_1$	$\vec{\mathbf{X}}'_2$
$\vec{\mathbf{X}}_1$	α	β
$\vec{\mathbf{X}}_2$	γ	κ

Fig. 10. Confusion Matrix for **Q**

4.5 Certainty States

The "most certain" state achievable for the example HRR radar example presented here is the case of the observed deterministic multivariate signal in noise ($\vec{\mathbf{X}}_n^i$) when accompanied by perfect training ($\vec{\mathbf{X}}' = \vec{\mathbf{X}}_n^i$). Assuming sufficient sampling to completely determine the *pdf* associated with the additive noise, the resulting statistical characteristics of the random performance functions will resemble the delta function [44], and thus the reliability in predicted link performance (such as $\sigma_{\mathbf{P}_e}$) will be very high as shown in case 1 of Fig. 10. In a less certain case, the *signal* under measurement is random in nature ($\vec{\mathbf{X}}$). The expected performance of the random performance functions will reflect the loss in information due to the degree of uncertainty present in $\vec{\mathbf{X}}$ as well as a decrease in reliability. Given the progressively large number of degrees of freedom associated with the uncertainty parameters associated with $\vec{V}_E$ and $\vec{V}_t$ in $\vec{\mathbf{X}}$, the statistical support underlying the statistics of the random link performance functions $\mathbf{P}_\mathbf{e}^{\vec{\mathbf{X}}}, \mathbf{P}_\mathbf{e}^{\vec{\mathbf{Y}}}, \mathbf{P}_\mathbf{e}^{\mathbf{Q}}$, and $\mathbf{P}_\mathbf{e}^{\mathbf{Q}'}$ can quickly increase as is shown in case 3-5 below in Fig. 11.

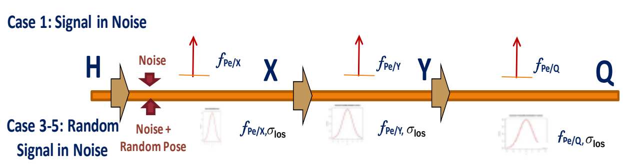


Fig. 11. Propagation of Uncertainty Illustration

Table IV below relates selected combinations of measurement and training uncertainty sources from Table 1. The cases 1-6 identified in Table IV represent the certainty states of interest within the system. Case 1 of Table IV represents an observed process $\vec{\mathbf{X}}_n$ of a stationary object of known aspect angle with perfect training. Case 1 conditions correspond to the *highest certainty state possible.* Case 2 corresponds to the observed process $\vec{\mathbf{X}}$ of an object that is moving slow enough as to appear stationary during the measurement interval. The aspect estimation is $\sigma_\mathbf{t}$ =.75

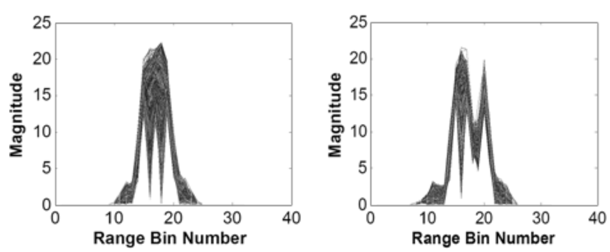
degrees with an unknown bias (μ_t) and again the training is perfect. Case 3 conditions are similar with an unknown leading edge position bias μ_r .

The parameter SNR is treated as an unknown parameter in Case 4. Case 5 is a combined condition of the unknown parameters in Case 2, 3, and 4. In case 6, a form of imperfect training is presented where the measurement parameter uncertainty provided in Case 5 is combined with training level B (μ_r = 0 and μ_t = 0).

4.6 Sampling and FBIT Analysis

4.6.1 Signature Ensembles

The amplitude response for the N sample ensemble of HRR signatures for a "baseline" set of conditions defined as Case 2 (μ_r = 0 and μ_t = 0) are provided in Fig. 12.a and Fig. 12.b.



a. Magnitude of $\vec{x}_1$ b. Magnitude of $\vec{x}_2$

Fig. 12. HRR Signature Amplitude Ensemble (No Noise); N=10^3

TABLE IV

MEASUREMENT AND TRAINING CERTAINTY CASES

Case Number	**Training Level**	**Measurement Level**
Case 1	$\vec{X}' = \vec{X}_n$ $\sigma_t = 0^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB	$\vec{X}_n$ $\sigma_t = 0^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB
Case 2	$\vec{X}' = \vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB	$\vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB
Case 3	$\vec{X}' = \vec{X}$ $\sigma_t = .75^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \boldsymbol{\mu_r}$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB	$\vec{X}$ $\sigma_t = .75^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \boldsymbol{\mu_r}$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB
Case 4	$\vec{X}' = \vec{X}$ $\sigma_t = .75^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ ***SNR***	$\vec{X}$ $\sigma_t = .75^\circ, \mu_t = 0^\circ$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ ***SNR***
Case 5	$\vec{X}' = \vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \boldsymbol{\mu_r}$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ ***SNR***	$\vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \boldsymbol{\mu_r}$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ ***SNR***
Case 6	$\vec{X}' \neq \vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \mu_r = 0^\circ$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ SNR=20dB	$\vec{X}$ $\sigma_t = .75^\circ, \boldsymbol{\mu_t}$ $\sigma_r = 0m, \boldsymbol{\mu_r}$ $(\theta_t, \phi_t) = 10^\circ, 7.5^\circ$ ***SNR***

*note: Parameter $\boldsymbol{\mu_t}$ modeled uniform [$-1^\circ, 1^\circ$], Parameter $\boldsymbol{\mu_r}$ modeled uniform [$0^\circ, .2^\circ$], Parameter ***SNR*** modeled Gaussian ($\mu = 20$ dB, $\sigma^2 = 4$ dB)

The five target features (K=5) at range bins 17-21 are selected for discriminate processing in $\vec{X} \rightarrow \vec{Y}$.

4.6.2 Sampling Uncertainty Example

The sampling uncertainty defined in Section II.J is illustrated using the baseline uncertainty conditions and multiple target ensembles similar to those given above in Section III.I.2. Using the Monte-Carlo simulation, the typical sets for $\vec{X}_1$, $\vec{X}_2$, and $\vec{X}$, are computed for an increasing value for N. Multiple ensembles of each are simulated at each value of N to generate both the mean and variance of the entropy estimate within the typical set.

The typical set plot in Fig. 13 provides the value for N_M for the entropy estimates for $\vec{X}$ as defined in Section III.I.2. The number of samples required for each ensemble based on the phase transition at $N_T=2x10^3$ within the typical set profile is determined to be $N_M=2x10^5$.

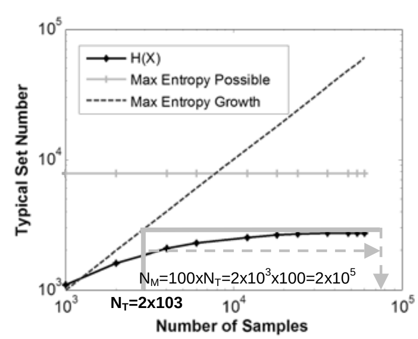


Fig. 13. Phase Transition Within Typical Set of $\vec{X}$ Versus N; n_b=6

Figure 14 demonstrates the entropy scaling property discussed in section III.J. In the following example, Monte-Carlo simulation is used to compute the actual estimation variance (L draws=1000) at each incremental setting of N_M. The estimation variance at $N_T=3x10^3$ is scaled to each setting of N_M to a maximum value of $N_M=2x10^5$ validating the use of the 1/N scaling factor in (47).

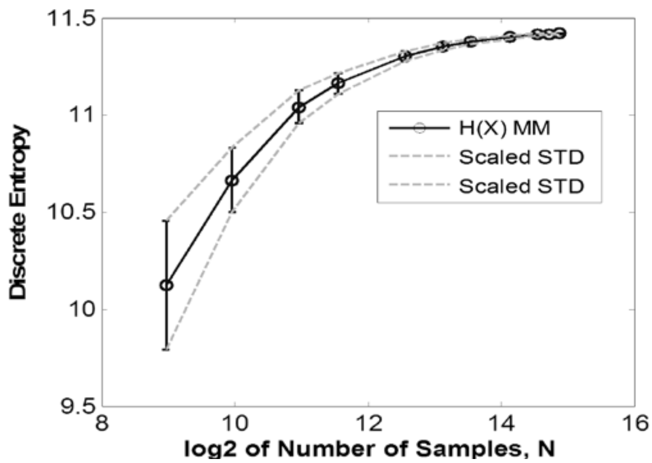


Fig. 14. Scaled Standard Deviation of Estimator of Entropy of $\vec{X}$ Versus N; $n_b=6$, L=1000

The sampling uncertainty associated with entropic estimation at $\vec{X}$ is realized within the estimate $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$. Figure 15 applies the 1/N scaling directly to the MLE estimate of $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$, $\hat{I}(\mathbf{H};\vec{\mathbf{Y}})$, and $\hat{I}(\mathbf{H};\mathbf{Q})$ beginning at $2xN_T = 6x10^3$.

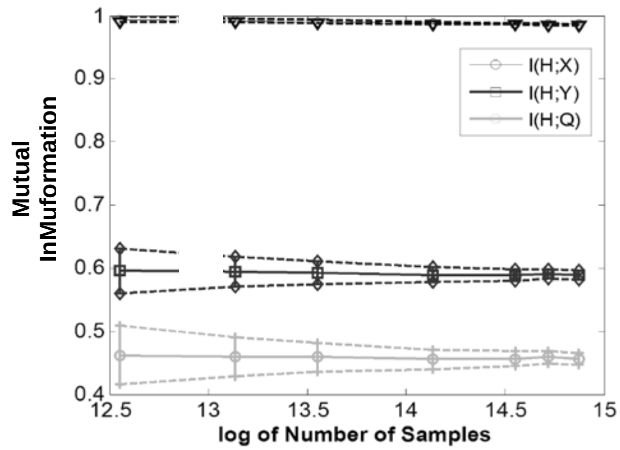


Fig. 15. Scaling Properties of $\hat{I}(\mathbf{H};\vec{\mathbf{X}})$, $\hat{I}(\mathbf{H};\vec{\mathbf{Y}})$, $\hat{I}(\mathbf{H};\mathbf{Q})$ Versus Ensemble Size N, N_T=3x10^3, n_b=6, L=1000

In Equation (47), Corollary IV and V are used to compute the sampling uncertainty associated with the estimate of the probability of error. The following figures demonstrate the accuracy of using (44). Equation (44) is applied at each link in the radar channel. Note that each application of (44) is conducted with the 2xN_T=6x10^3 as the basis for the scaling. The approximation for the standard deviation of the probability of error is computed for the complete range of ensemble size out to N=3x10^4. Figure 16 provides a comparison of the probability of error estimate using (44) to the error computed using simulation. These results show that the estimates compare very nicely to the “actual” results. This agreement indicates that the dispersion of the mean mutual information of the estimate is low enough to support the use of the linear approximation.

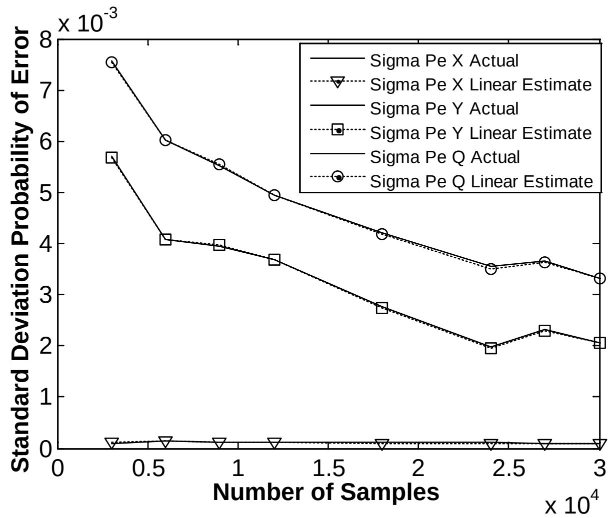


Fig. 16. $\sigma_{\hat{P}_{e_N}^{\tilde{X}}}$, $\sigma_{\hat{P}_{e_N}^{\tilde{Y}}}$, and $\sigma_{\hat{P}_{e_N}^{Q}}$ Versus the Linear Approximation, $N_T=3\text{x}10^3$, $n_b=6$, L=1000

The application of (44) at each draw of the Monte-Carlo simulation will generate an estimate of the sampling uncertainty associated with the probability of error estimate. Figure 17 illustrates the application of (44) to the results in Fig. 15.

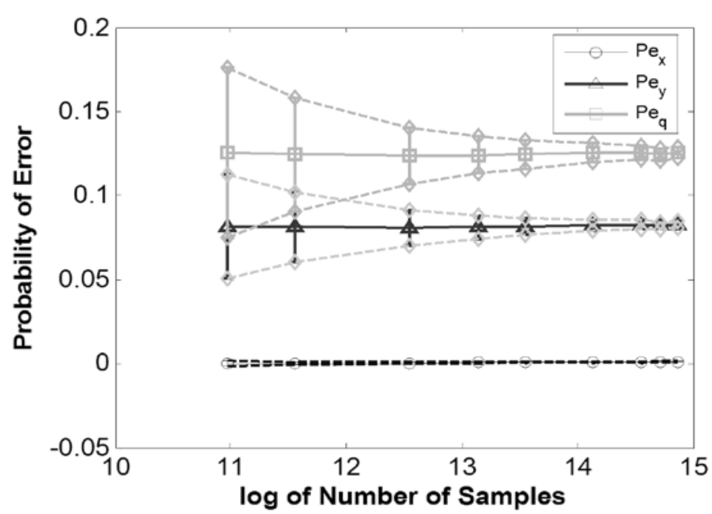


Fig. 17. Scaling Properties of $\hat{P}_e^{X}, \hat{P}_e^{Y}, \hat{P}_e^{Q}$ Versus N, $N_T=3\text{x}10^3$, L=1000

Equation (48) provides the test for minimum sampling based on low sample ensemble sizes. In Fig. 18, (48) is applied to the radar example at the three link positions $\vec{\mathbf{X}}, \vec{\mathbf{Y}}$ and $\mathbf{Q}$. The test results in Fig. 18 show that the true ratio of sampling variance to the variability in predicted link performance is given as a function of ensemble size N. This is indicated by the solid lines. The dashed lines represent the ratio as given by the 1/N scaling as discussed above. The required ratio α is given by the dashed black line at two different levels. The results of the test given in (48) are given at each increment for $N_T = 3\text{x}10^3$. *The interesting observation in* Fig. 18 *is that the point at which the test falls below the threshold* α *is consistent with the ensemble size* N_M *as derived from the phase transition point* N_T *as outlined in section* III.K. This is a significant validation of the use of the phase transition method for estimating minimum ensemble size within Monte-Carlo simulation. The results of the three tests above provide insight into the relationship of the required ensemble size N to the reliability in link performance estimates within sensitivity analysis simulations.

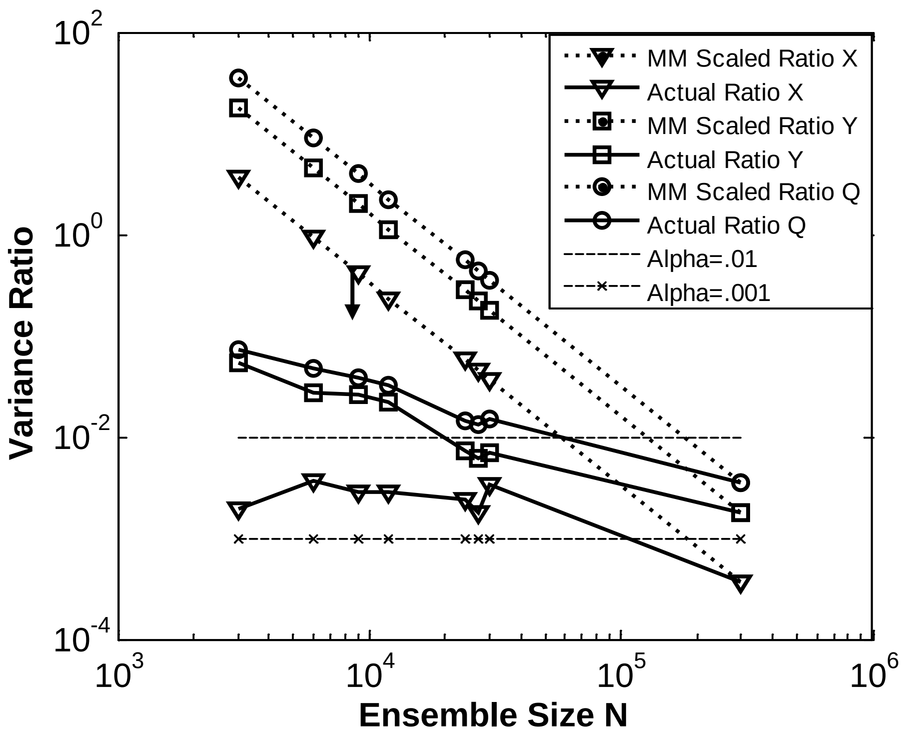


Fig. 18. Sampling Uncertainty and Testing N_M at $\vec{\mathbf{X}}, \vec{\mathbf{Y}}, \mathbf{Q}$

4.6.3 The Fano Equality

It is important to demonstrate the validity of Theorem I as written in (2). Using the radar example, Fig. 19 illustrates that the addition of I(**Q;V**) brings the approximation form of Fano into agreement with the "true" probability of error as simulated using Monte-Carlo within the radar example outlined above. Again using Case 2 (μ_r = 0 and μ_t = 0) conditions for the binary classification, the performance given by the Fano approximation is given by the red line. The Simulated "true" (actual) performance is given by the black line. The green line represents the performance using the equality (exact) form of Fano in Theorem I. The equality form of Fano agrees with the "true" performance which validates Theorem I.

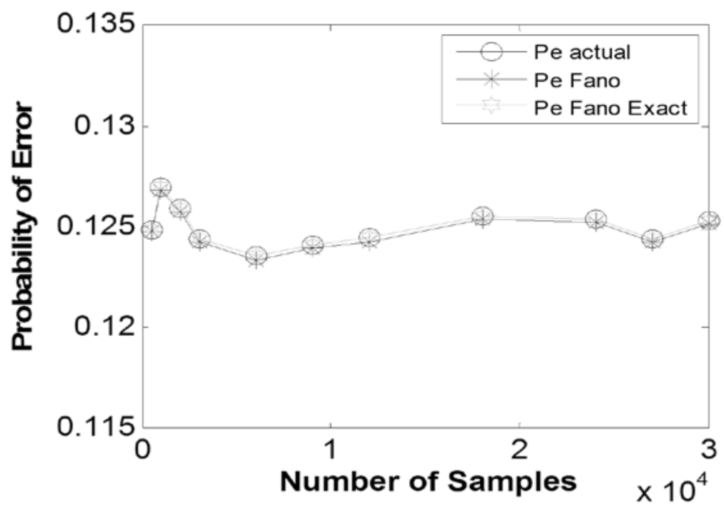


Fig.19. Actual Mean Probability of Error at **Q** Versus Mean Fano Estimated Performance at **Q,** n_b=6, L=1000

5 EXPERIMENTS

Several hypotheses directly highlight the potential advantages of the FBIT method in the performance characterization of an information sensing system in the presence of various uncertainties.

5.2 Research Hypotheses

Hypothesis 1: Information flow through the components of a sensing system can be studied and information bottlenecks can be identified. System performance upper bounds can be characterized based on the loss in information attributed to each component.

Hypothesis 2: For a fixed H(**H**), maximizing I(**H,Q**) will minimize the equivocation H(**H/Q**) and thus minimize the probability of error P_e. Thus; system component design parameters (ν_c) can be traded directly with loss in the channel *to minimize* P_e.

Hypothesis 3: Sources of uncertainty can be characterized in terms of their effects within the decision rule subspaces and their relative impact to losses within several subsystem components of the radar system.

Hypothesis 4: The increasing dimensionality of $\vec{\mathbf{x}}$ will eventually lead to an unacceptable degradation to the reliability in predicted link performance.

Hypothesis 5: Selected sources of uncertainty within the radar information channel can be ranked as to their relative impact to the performance of the information exploitation system.

5.3 Experiments

The experiments conducted to address the hypotheses above are given in Table V.

TABLE V.

LIST OF EXPERIMENTS AND APPLICABLE CASES

Experiment	Case	Hypothesis
1. Information Flow	2	1
2. System Trades	2	2
3 System Uncertainty and Information Flow	1,2,3,4,5,6	3,4,5

5.3.1 Information Flow and Design Trades within the Radar Channel

The value of the Data Processing Inequality is readily seen from Fig. 20-22 below where the individual loss at each link in the channel can be quantified. In each of the figures, the MI and probability of error is computed for a changing design parameter within ν_c. Three design parameters are traded; system thermal noise, system dynamic range, and system bandwidth.

The signal-to-noise ratio of the signatures resulting from sensor measurements depend in part on the noise figure of the system. In Fig. 20 thermal noise is scaled by varying the noise figure across a range that affects a SNR range of 1 dB to 10 dB (SNR is given in frequency domain prior to inverse Fourier Transform gain). The results of the SNR trade indicate that an SNR of 8 dB in the frequency domain (19 dB in the time-delay domain after transform gain) will generate maximum information flow.

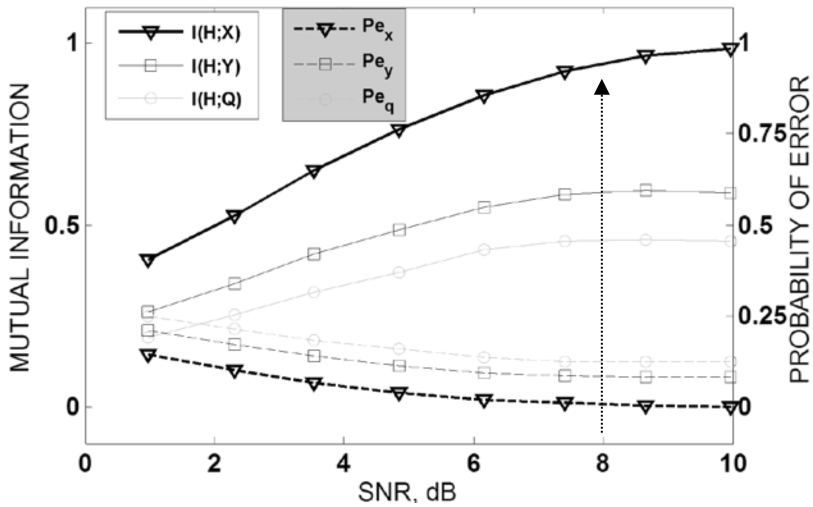


Fig.20. System Thermal Noise Trade, SNR in Frequency Domain

It is also of interest how the dynamic range of the sensor affects the information flow through the channel. Specifically, the sensitivity of $I(H;Q)$ and ultimately P_e to the dynamic range in the sensor is of interest. The A/D conversion of the radar intermediate frequency (IF) signal to a digital representation must preserve the amplitude and phase information contained in the radar return with minimum error. The effects of quantization at each measurement point (quantization event) due to the twos-complement rounding error are assumed to be zero mean white noise processes [45]. The A/D conversion and associated quantization noise are modeled as an additive noise component $\vec{e}$ and added to the measured signature process [46].

$$\vec{X}_n^i = \vec{X}_E^i + \vec{n} + \vec{e} \tag{52}$$

The maximum dynamic range supportable by a “B-bit” quantizer is the ratio of the largest representable magnitude to the smallest nonzero representable magnitude [47]. The dynamic range for twos compliment and magnitude encoding for a “B-bit” quantizer is [48]

$$\text{Dynamic Range (dB)} = 20 \cdot \log_{10}\left(\frac{2^{(B-1)} - 1}{1}\right).$$

The trade in Fig. 21 indicates that a 3 or 4 Bit A/D converter is needed to maximize information flow in the channel given the binary target set under evaluation.

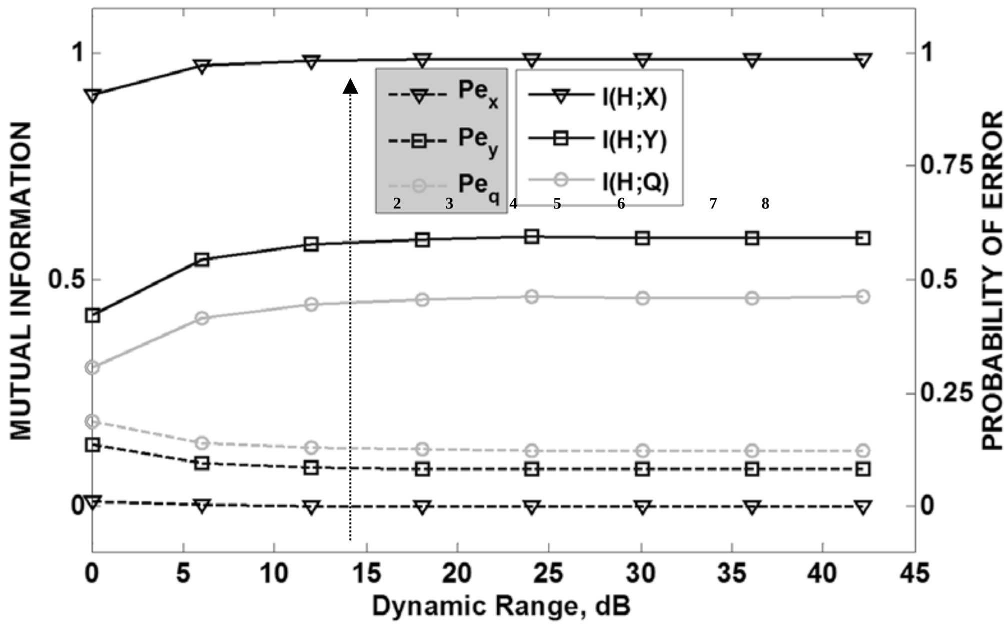


Fig. 21. System Dynamic Range Trade

The analysis of the bandwidth trade in Fig. 22 can be nicely linked to the physical scattering configurations of target 1 and target 2. As was mentioned earlier in the report, the locations for the non-collocated dominant scatterer differ by .2 meters or .65 feet.

One would then expect that there should be a ‘bump’ in information flow when the bandwidth reaches levels that support the resolution necessary to resolve the peaks associated with these two scatterers. The theoretical resolution to achieve this feature separation would be approximately 800 MHz using the fundamental bandwidth relationship;

$$\frac{c \cdot \tau}{2} = .6\ feet = \frac{c}{2 \cdot BW}.$$

In Fig. 22 the bump in performance is centered at 800 MHz where the mutual information at $\vec{Y}$ and $\mathbf{Q}$ is rapidly increasing and where the probability of error is greatly reduced.

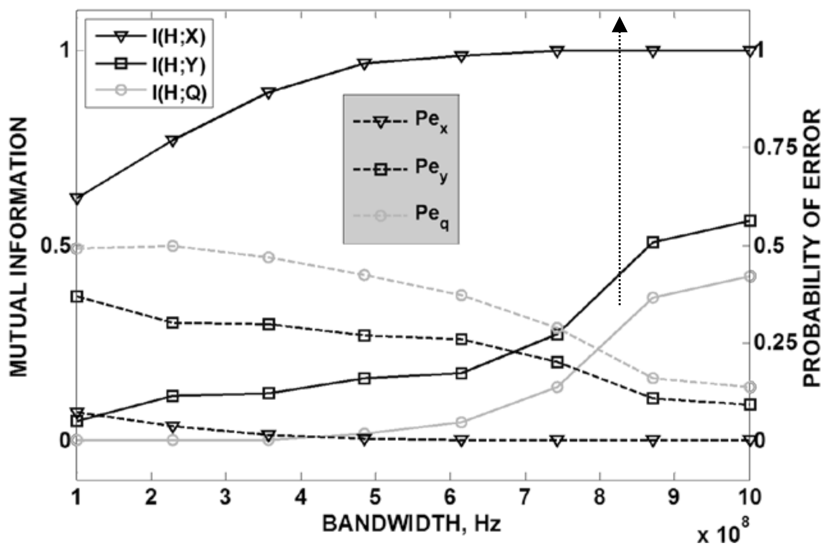


Fig. 22. System Bandwidth Trade

In each figure it can be seen that the MI decreases as links move further down the channel. With one Bit going into the channel (binary classification problem), Table VI below tabulates the information loss budget for each trade study at the selected baseline operating point.

The study of Table VI reveals several key points. First, In this particular example problem, the targets appear to be separating very well at $\vec{X}$, and much of the loss occurs within the feature extraction and at the application of the decision rule. The loss at link $\vec{Y}$ appears to be the dominant information limiting component in the system. There is a loss of .3-.4 Bits at the feature extraction function at $\vec{Y}$. The information loss associated with signature measurement and signature processing results in only .1 Bits of loss. This is very important information in achieving optimization of the system design for information sensing. *Little gain can be expected through the expansion of sensing degrees of freedom (DOF) in improving the overall performance of the system.*

TABLE VI
INFORMATION LOSS BUDGET FOR VARIOUS TRADES

System Component	**Information Loss, Bits**		
	Trade 1 SNR	**Trade 2 DR**	**Trade 3 BW**
Source-to-Measurement ($\vec{\mathbf{X}}$)	0.1	0.1	0.05
Measurement –to-Feature ($\vec{\mathbf{Y}}$)	0.4	0.3	0.4
Decision Rule Application ($\mathbf{Q}$)	0.1	0.2	0.1
Total Channel Loss*	0.6	0.6	0.55

*Baseline Conditions; SNR=20 dB, BW=800 MHz, DR=20 dB

Also, the loss due to the decision component of the system is in the range of .1 - .2 Bits. Depending on the performance requirements of the system, improvements to the decision stage of the system may or may not be warranted. Prior to the decision stage of the system, .4 - .5 bits of cumulative loss have been sustained resulting in an "upper bound" in performance of something in the area of P_e =.1. *No improvements to the classifier design within the decision component of the system can improve upon this performance level*. Improvements appear to be best directed toward the feature extraction stage.

An optimal design operating point may for example include the following component selections; (i) A/D converter with B=4 Bits, (ii) Receiver design which achieves 20 dB SNR under tactically significant conditions, and (iii) Transmit waveform with BW> 800 MHz.

5.3.2 Information Flow and System Uncertainty

The study of the effects of sources of uncertainty on system performance confidence while under control parameters $\vec{v}_c$ and in the presence of sensing uncertainty ($\vec{V}_E, \vec{V}_t$) is of particular interest. For a fully sampled signature process with negligible sampling uncertainty per (46), the FBIT method can be applied to study the independent sources of uncertainty. The effects of each independent source of uncertainty can be studied at each link in the channel. Equation (36) is demonstrated for links $\vec{X}$, $\vec{Y}$, and Q under case 5 conditions defined in Table IV. Under these conditions, three independent sources of uncertainty are introduced in the system under perfect training conditions. An unknown bias in target aspect estimation and an unknown bias in leading edge range bias estimation are assumed. The target range is also unknown and as such a third uncertainty in introduced in the SNR of the measured signature. All assumed statistics associated with the uncertainties are as defined under case 5 of Table IV and as described in section IV.D.

Using Monte-Carlo simulation L independent draws of an N_M sample ensemble from $\vec{x}$ are generated. The FBIT method is applied at each draw to generate the decomposition of the performance estimate reliability in (36) at $\vec{X}$, $\vec{Y}$, and Q. In Fig. 23 the cumulative link loss standard deviation defined in (24.a), (24.b), and (24.c) resulting from the sum of the independent three uncertainty sources is computed about the expected link information loss defined in (23.a), (23.b), and (23.c). To clearly illustrate the level of agreement of the independent link loss contributions to the total produced by the joint simulation, the individual contributions to the cumulative link loss variance are individually plotted in an incremental fashion in Fig. 23. Fig. 23 shows that the sum of the independent uncertainty sources yields the same results as the Monte-Carlo simulation involving all three factors in a joint process.

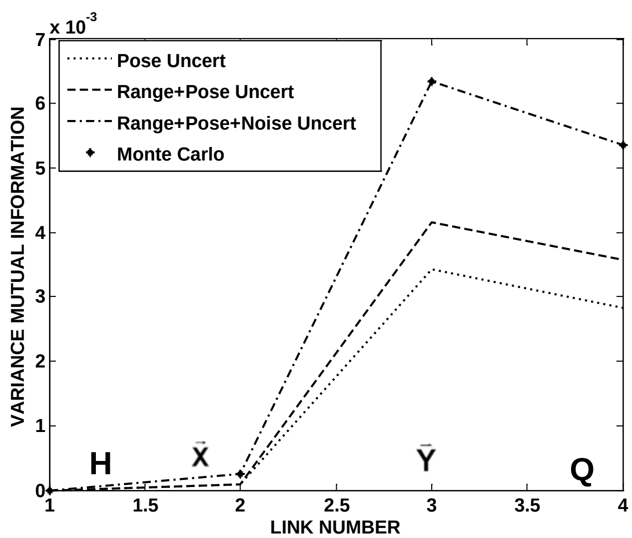


Fig. 23. Cumulative Channel Link Loss Variance, L=100, N=3*10^5

The corresponding impacts to the reliability in link performance can be generated through the application of Corollary IV and V. The reliability in predicted link performance as quantified by Definition VI resulting from the sum of the independent three uncertainty sources is presented in the error bars about the expected link performance defined in (31.a), (31.b), and (31.c). The dashed line represents the results of the joint Monte-Carlo simulation where all three independent uncertainty factors are simulated simultaneously. The results in Fig. 24 show clearly that the sum of the independent events equals the joint event, thus validating the assumption of independence in the three sources of uncertainty acting on the predicted performance risk.

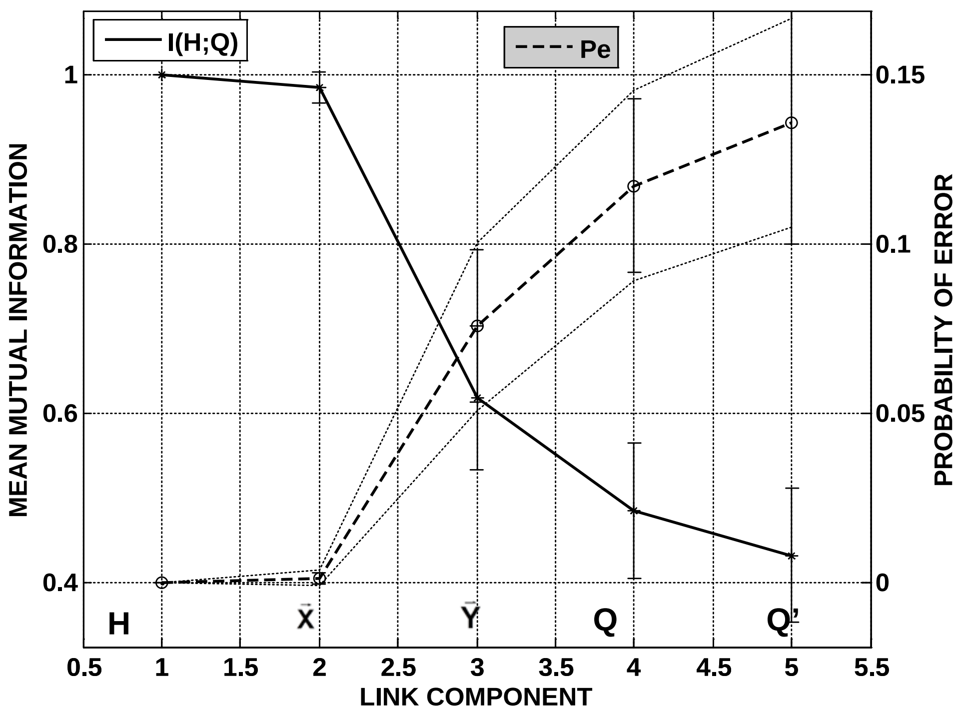


Fig. 24. Channel Mutual Information and Reliability in Link Performance, L=100, N=3*10^5

In Fig. 25 a similar validation of the propagation of independent uncertainty sources is given for the reliability in predicted performance. The example demonstrates that the use of Corollary IV and V to approximate the reliability on the performance estimate using the link loss variance is a very effective means to address the transcendental relationship underlying this method. The data points marked with the asterisks represent the sum of the independent contributions to the reliability in performance prediction. The respective plotted lines represent the results of direct simulation at the specified link.

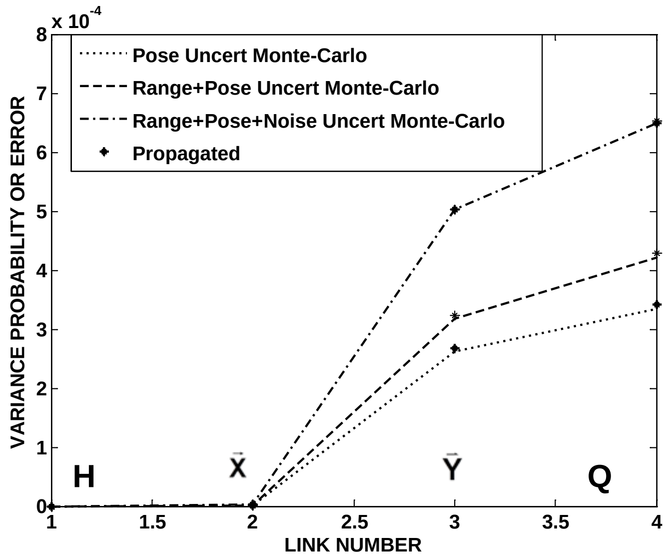


Fig. 25. Reliability in Predicted Link Performance Given as Variance, L=50, N=3*105

The implications of imperfect training are realized in the final stage of the channel at **Q**' as shown in Fig. 24. At **Q**', Case 6 conditions in Table IV are used to present a naive training approach as developed in section IV.D.

A summary of the expected link loss, expected link performance, reliability in link performance, and results of respective sampling uncertainty tests in Fig. 18 are given in Table VII below. The reliability in predicted performance decreases as information propagates down the sensing channel. The expected link performance also decreases in accordance with the principles of mutual information and the Data Processing Inequality. Much of the decrease in reliability and loss in predicted performance and loss in performance comes at the feature extraction stage in the system. The reduced reliability in performance prediction is most sensitive to the uncertainty factor of SNR. The effects of the factors associated with target range bias and pose estimate bias are of less significance relative to the total reliability in predicted performance.

From Table VII it can be seen that gains in performance due to component design trades must also take into account the reliability level associated with predicted performance.

TABLE VII
INFORMATION CONFIDENCE & LOSS BUDGET FOR VARIOUS

Link	Link Information Measure			
	Link Loss, Bits	Expected Link Performance	Reliability in Link Performance	Sampling Uncertainty Test
H	0.0	--------	----	------
$\vec{\mathbf{X}}$	$\mu_{\mathbf{IL}_{S\Delta}} = .05$	$\mu_{\mathbf{P}_{e_{\vec{X}}}} = .013$	$\sigma_{P_e^{\vec{X}}} = .003$	$\left(\frac{\sigma^2_{\hat{P}^{\vec{X}}_{eN}}}{\sigma^2_{P^{\vec{X}}_{eN}}}\right) < .001$
$\vec{\mathbf{Y}}$	$\mu_{\mathbf{IL}_{F\Delta}} = .35$	$\mu_{\mathbf{P}_{e_{\vec{Y}}}} = .073$	$\sigma_{P_e^{\vec{Y}}} = .0228$	$\left(\frac{\sigma^2_{\hat{P}^{\vec{Y}}_{eN}}}{\sigma^2_{P^{\vec{Y}}_{eN}}}\right) < .003$
Q	$\mu_{\mathbf{IL}_{D\Delta}} = .16$	$\mu_{\mathbf{P}_{e_Q}} = .12$	$\sigma_{P_e^Q} = .0255$	$\left(\frac{\sigma^2_{\hat{P}^{Q}_{eN}}}{\sigma^2_{P^{Q}_{eN}}}\right) < .006$
Q'	$\mu_{\mathbf{IL}_{T\Delta}} = .04$	$\mu_{P_{e_{Q'}}} = .125$	$\sigma_{P_e^{Q'}} = .0266$	---

In this example problem, changes within two significant digits of the expected performance should be studied in the context of the reliability of the performance estimates based on uncertainty factors introduced in the system.

6 CONCLUSION

The FBIT method is developed for use in the research of information sensing applications. Measures are developed to identify information flow bottlenecks and to form an information link budget for system analysis. Techniques for bounding asymptotic performance under sufficient sampling are characterized. Test criteria are developed for controlling sampling uncertainty within the uncertainty propagation analysis approach. Test criteria are linked to phase transitions within the typical set trajectory associated with the entropy estimation of high dimensional signature processes. The FBIT method and test criteria are applied to an HRR radar numerical example. The propagating effects of various sensing uncertainties on system performance are characterized at the component level.

7 REFERENCES

[1] R. Ahlswede, N. Cai, S. Robert, and R. Yeung "Network Information Flow," *IEEE Transactions on Information Theory* Vol. 46, no. 4, July. 2000.

[2] T. Cover, J. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[3] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[4] J. A. Cortese, Quantum Information Science and Superconducting Device Technology, MIT Lincoln Laboratory Technical Memorandums to AFRL/RYAS, 2011-2012, Unpublished.

[5] R. Calderbank, S. D. Howard, and B. Moran, "Waveform Diversity in Radar Signal Processing, A Focus on the Use and Control of Degrees of Freedom", *IEEE Signal Processing Magazine* (32), January 2009.

[6] P.M. Woodward, I.L Davies, "A Theory of Radar Information", *Philosophical Magazine*, Vol. 41, pp. 1001-1017, October 1951.

[7] P.M. Woodward, Probability & Information Theory with Applications to Radar, 2nd edition, Pergamon Press, 1964

[8] C. E. Shannon, `A mathematical theory of communication,' *Bell System Technical Journal*, Vol. 27, 379-423 and 623-656, July and October, 1948.

[9] M. Bell, "Information Theory and Radar: Mutual Information and the Design and Analysis of Radar Waveforms and Systems", Ph.D. Dissertation, Department of Electrical Engineering, California Institute of Technology, 1988.

[10] A. Leshem, O. Naparstek, A. Nehorai, "Information Theoretic Adaptive Radar Waveform Design for Multiple Extended Targets", *IEEE Transactions on Selected Topics in Signal Processing*, Vol. 1, No. 1, June 2007.

[11] S. Sowelam, A. Tewfik, "Waveform Selection in Radar Target Classification", *IEEE Transactions on Information Theory*, Vol. 46, no. 3, May 2000.

[12] S. Kullback, R.A. Leibler, "On Information & Sufficiency", *Ann. Math. Stat*., 22: 79-86, 1951.

[13] S. Briles, "The Theory for, and the Demonstration of, Information Theory Applied to Radar Target Identification", Ph.D. Dissertation Kansas State University, 1992.

[14] M Horne "An Information Theory for the Prediction of SAR Target Classification Performance" DERA, Malvern, *WR14 3PS, UK2, SPIE,* Vol. 4382, 2001, Algorithms for Synthetic Aperture Radar Imagery VIII.

[15] J.C. Principe, D. Xu, Q. Zhao, J.W. Fisher "Learning form Examples with Information Theoretic Criteria", Journal of VLSI Signal Processing 26, 61–77, 2000.

[16] J. O'Sullivan, M. DeVove, V. Kedia, M. Miller, "SAR ATR Performance Using a Conditionally Gaussian Model", IEEE *Transactions on Aerospace Electronics Systems Vol.37, No. 1, Jan. 2001.*

[17] K. Pasala, J. Malas, "HRR Radar Signature Database Validation for ATR; An Information Theoretic Approach", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 47, No. 2 April 2011.

[18] E. Fishler, A. Haimovich, R. Blum, L. J. Cimini, D. Chizhik, and R. A. Valenuela, "Spatial Diversity in Radars-models and Detection Performance," *IEEE Transactions on Signal Processing*, Vol. 3, no. 54, pp. 823-838, Mar. 2006.

[19] K. Forsythe and D. Bliss, "Waveform Correlation and Optimization Issues for MIMO Radar," in *Proc. 39th Asilomar Conf. Signals, Systems, and Computers*, Monterey, CA, 2005, pp. 1306-1310.

[20] D. R. Fuhrmann and G. San Antonio. "Transmit Beamforming for MIMO Radar Systems Using Partial Signal Correlation," *IEEE Transactions Aerospace Electronics Systems* Vol. 43, no. 4, Oct. 2007.

[21] D. Bliss and K. Forsythe, "Multiple-input Multiple-output (MIMO) Radar Imaging: Degrees of Freedom and Resolution," in *Proc. 39th Asilomar Conf. Signals, Systems, and Computers*, Monterey, CA, 2003, pp. 54-59.

[22] N. Tishby, F. Pereira, and W. Bialek. "The information bottleneck method" in *Proc. 37th Allerton Conf. on Communication and Computation*, 1999.

[23] C. H. Papadimitriou, K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* "6.1 The Max-Flow, Min-Cut Theorem", pp. 120–128, Dover., 1998

[24] B. C. Geiger, C. Feldbauer, G. Kubin, "Information Loss in Static Nonlinearities" , in Proc. IEEE Int. Sym. *Wireless Communication Systems (ISWSC)*, Aachen, Nov. 2011, accepted; preprint available: arXiv:1102.4794 [cs.IT].

[25] N. Merhav,"Data processing inequalities based on a certain structured class of information measures with application to estimation theory" *IEEE Trans. Inform. Theory,* vol. 58, no. 8, pp. 5287-5301, August 2012.

[26] T. Santner, B. Williams, W. Notz, *The Design and Analysis of Computer Experiments*, Springer, 2003.

[27] B. Ayyub, G. Klir, *Uncertainty Modeling and Analysis in Engineering and the Sciences*, Chapman & Hall/CRC, 2006.

[28] H. Coleman, W. Steele, *Experimentation Validation, and Uncertainty Analysis, for Engineers*, 3rd Edition, John Wiley & Sons 2009.

[29] J. A. Sokolowski, C. M. Banks, *Modeling and Simulation Fundementals: Theoretical Underpinnnings and Pratical Domains*, John Wiley & Sons, 2010.

[30] A. Saltelli, et al, *Global Sensitivitoes Analysis, The Primer*, England: Wiley, 2008.

[31] J. Malas, J. Cortese, "The Radar Information Channel and System Uncertainty", *IEEE proceedings to the 2010 IEEE Radar Conference, Washington DC.*

[32] J. Mandel, *The Statistical Analysis of Experimental Data*, National Bureau of Standards, Washington*, D.C.,* Dover Publications Inc.

[33] J. A. Cortese, "Information Thoery and Radar Signal Processing", Air Force Research Lab Technical Report, Georgia Institute of Technology, College of Computing, March, 2009, Unpublished

[34] S. Geman et al., "Neural Networks and the Bias/Variance Dilemma", Journal of Neural Computation Vol. 4, pp. 1-58, 1992.

[35] R. J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, Reading MA, 1977.

[36] L. Paninski, "Estimation of Entropy and Mutual Information", Journal of Neural Computation, Vol. 15, pages 1191-1253, 2003, MIT Press.

[37] G. Miller, W. Madow, , *On the Maximum Likelihood Estimate of the Shannon-Wiener Measure of Information*, Technical Report AFCRC-TR-54-75, August 1954, Operational Applications Laboratory, Air Force Cambridge Research Center, Air Research and Development Command Bolling Air Force Base, Washington 25, D.C.

[38] T. Hogg, "Phase transitions and the search problem by, artificial intellience", (an Elsevier journal) volume 81, published in 1996, Pages 1-15.

[39] R. Monasson, R. Zecchina, "Statistical mechanics of the random k-satisfiability model", phy. rev. e (pre) volume 56, number 2, page 1357, August, 1997.

[40] S. Mertens, "Phase transition in the number partitioning problem", prl, vol 81, no 20, Nov 16, 1998, Page 4281.

[41] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman and L. Troyansky, "Determining computational complexity from characteristic phase transitions" by, *Nature* (London) vol. 400, July 8, 1999.
[42] B. Bollobas, C. Borges, J. Chayes, J. Kim, D. Wilson, "The scaling window of the 2-sat transition", *Random Structure and Algorithms* (a Wiley journal), volume 18, page 20, 2001.
[43] N. Levanon, *Radar Principles*, Wiley, New York, 1988.
[44] H. L. Vantrees, *Detection, Estimation, and Modulation Theory*, Part I, , John Wiley & Sons, 2001.
[45] A. Oppenheim, R. Schafer, *Discrete-Time Signal Processing*, Second Edition, Prentice Hall, 1999.
[46] J. Scheer, J. Kurtz, *Coherent Radar Performance Estimation*, Artec House, 1993.
[47] M. Richards, *Fundamentals of Radar Signal Processing*, McGraw Hill, 2005.
[48] A. Oppenheim, Editor, *Applications of digital Signal Processing*, Prentice Hall, 1978, page 663.
[49] Document approved for public release:88ABW-2013-0274, Jan 23, 2013

8 AUTHOR BIBLIOGRAPHIES

Dr. John Malas serves as Principle Research Scientist in the Layered Sensing and Exploitation Division of the Sensors Directorate within AFRL. Since 1983, Dr. Malas has worked in the areas of radar systems, target scattering, and signature information sensing supporting the research and development of technology for USAF airborne sensing systems. He received the Ph.D. degree in Electrical Engineering from the University of Dayton in 2002, M.S. and B.S. degree in Electrical Systems Engineering in 1990 and 1983 respectively from Wright State University. Dr. Malas is a registered Professional Engineer in the state of Ohio, an adjunct professor at the University of Dayton, and a senior member of IEEE.

Ms. Ryan is the Technical Advisor of the Radio Frequency (RF) Exploitation Branch of the Sensors' Directorate Layered Sensing Exploitation Technology Division. Ms. Ryan has twenty five years of experience improving radar system and performance across DoD, NASA and FAA applications. Her expertise includes flight test planning and execution and data analysis/simulation for radar design trades. She has extensive experience in the area of intelligence production in support of advanced radar technology. Ms. Ryan's skills include high fidelity multi-sensor signature modeling and validation
and physics-based exploitation methods for Tri-service and Missile Defense Agency (MDA) applications. She is knowledgeable in radar cross section measurement, calibration and data analysis. Her breadth of experience extends to radar and multi-sensor track correlation algorithms and innovative radar waveforms for MDA applications and RCS based feature aided tracking for Ascent Debris. She has served as the Nation Representative for multiple radar focused NATO SET Panels 2009 - 2014. Ms. Ryan received the M.S. degree in Electrical Engineering from The Ohio State University in 1987 and the B.E.E. degree in Electrical Engineering in 1985 from The Georgia Institute of Technology. She was named a Senior Member of the IEEE in 2006.

Dr. John Cortese received an MS in Applied Mathematics, a PhD in Electrical Engineering and a PhD in Theoretical Physics from the California Institute of Technology. Dr. Cortese is employed by MIT Lincoln Laboratory. Dr. Cortese may be contacted via email at jcort@LL.mit.edu